

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:34:09 ; Search time 22.3488 Seconds
(without alignments)
1735.006 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 403
Sequence: 1 MLHHVKLIYATSKRLVGKK.....MKKRELARIWDEIEKKLS 403

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 283416 seqs, 96216763 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	17.9	401	2 G71018	pantothenate metab
2	60	14.9	401	2 E75114	dna/pantothenate m
3	12	3.0	394	2 A72223	pantothenate metab
4	11	2.7	403	2 A64414	pantothenate metab
5	10	2.5	394	2 D81819	hypothetical prote
6	10	2.5	399	2 A81058	DNA/pantothenate m
7	10	2.5	399	2 A11302	pantothenate metab
8	10	2.5	399	2 A11674	pantothenate metab
9	9	2.2	404	2 D69455	pantothenate metab
10	8	2.0	127	2 T23443	hypothetical prote
11	8	2.0	210	2 S62355	high mobility grou
12	8	2.0	265	2 A99249	hypothetical prote
13	8	2.0	298	2 B97226	dihydroorotate deh
14	8	2.0	391	2 E71125	probable aromatic-
15	8	2.0	396	2 T36548	probable transamin
16	8	2.0	434	2 AF1238	E. coli YbdN prote
17	8	2.0	434	2 AB1601	E. coli YbdN prote
18	8	2.0	480	1 TVMTGC	phosphoglycerate k
19	8	2.0	481	2 D71927	cag island protein
20	8	2.0	481	2 B64585	cag pathogenicity
21	8	2.0	794	2 T36972	probable membrane
22	8	2.0	1261	2 S75130	sensory transducti
23	7	1.7	36	2 T03348	gene e19 protein -
24	7	1.7	60	2 G69115	hypothetical prote
25	7	1.7	77	2 E97156	exonuclease VII sm
26	7	1.7	91	2 F69252	hypothetical prote
27	7	1.7	101	2 F90925	hypothetical prote
28	7	1.7	101	2 B85774	hypothetical prote
29	7	1.7	111	2 F70338	methionine-trNA li

30	7	1.7	124	2 C70537	hypothetical prote
31	7	1.7	130	2 F75075	hypothetical prote
32	7	1.7	138	2 C64382	hypothetical prote
33	7	1.7	140	2 E75056	probable translati
34	7	1.7	140	2 D71104	hypothetical prote
35	7	1.7	144	2 A96580	50S ribosomal prot
36	7	1.7	145	2 A97222	ribosomal protein
37	7	1.7	145	2 E71723	hypothetical prote
38	7	1.7	149	2 D97333	protein tyrosine p
39	7	1.7	160	2 T10278	ribosomal protein
40	7	1.7	170	2 T43779	disulfide bond for
41	7	1.7	173	2 G82818	flavoprotein (impo
42	7	1.7	183	2 H95142	N-terminal region
43	7	1.7	183	2 F98010	nonhistone chromos
44	7	1.7	186	2 B61611	hypothetical prote
45	7	1.7	190	2 C90260	hypothetical prote

ALIGNMENTS

RESULT 1

G71018
pantothenate metabolism flavoprotein dfp homolog PH1444 - Pyrococcus horikoshii
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C:Accession: G71018
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: G71018
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-401 <KAW>
A:Cross-references: UNIPROT:O59114; GB:AP000006; NID:g33236133; PIDN:BAA30551.1; PID:g33
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1444

Query Match 17.9%; Score 72; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 6.6e-63;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 TGNPVTITGTGFIHVELAGEHENKADLLVCPATANTISKIACGIDDPVTVTVTAPP 127
DB 69 TGNPVTITGTGFIHVELAGEHENKADLLVCPATANTISKIACGIDDPVTVTVTAPP 128

QY 128 HIPIMIAPAMHE 139
DB 129 HIPIMIAPAMHE 140

RESULT 2

E75114
dna/pantothenate metabolism flavoprotein (dfp) PAB1897 - Pyrococcus abyssi (strain Orsa
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: E75114
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A:Reference number: A75001
A:Accession: E75114
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <KAW>
A:Cross-references: UNIPROT:Q9V0S3; GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB496;
A:Experimental source: strain Orsay
C:Genetics:


```

.; Dominguez-Bernal, G.; Ducaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: A11302
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-399 <GLA>
A;Cross-references: UNIPROT:Q8V674; GB:NC_003210; PIDN:CAC99903.1; PID:g16411279; GSPDB:
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lml825
C;Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVTAGATRE 206
|||||
Db 190 VLVTAGATRE 199

RESULT 8
A11674
pantothenate metabolism flavoprotein homolog lin1939 [imported] - Listeria innoc
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: A11674
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Ducaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: A11674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-399 <GLA>
A;Cross-references: UNIPROT:Q92AI3; GB:AL592022; PIDN:CAC97169.1; PID:g16414440; GSPDB:G
A;Experimental source: strain Clp11262
C;Genetics:
A;Gene: lin1939
C;Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVTAGATRE 206
|||||
Db 190 VLVTAGATRE 199

RESULT 9
D69455
pantothenate metabolism flavoprotein dfp homolog AF1645 - Archaeoglobus fulgidus
NAlternate names: probable aspartate 1-decarboxylase activase
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: D69455
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woose, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

```

```

A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: D69455
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-404 <KLE>
A;Cross-references: UNIPROT:Q28628; GB:AE000989; GB:AE000782; NID:g269312; PIDN:AAB895
C;Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.2%; Score 9; DB 2; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 TANTISKIA 110
|||||
Db 103 TANTISKIA 111

RESULT 10
T23443
hypothetical protein K08C9.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T23443
R;Lennard, N.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19741
A;Accession: T23443
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-127 <WIL>
A;Cross-references: UNIPROT:Q9XU72; EMBL:Z81567; PIDN:CA804589.1; GSPDB:GN00019; CESP:
C;Experimental source: clone K08C9
C;Genetics:
A;Gene: CESP:K08C9.6
A;Map position: 1
A;Intons: 52/2, 69/1
C;Superfamily: Caenorhabditis elegans hypothetical protein K08C9.6

Query Match 2.0%; Score 8; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 FRIRIKL 255
|||||
Db 85 FRIRIKL 92

RESULT 11
S62355
high mobility group protein 1 - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: S62355
R;Nightingale, K.; Dimitrov, S.; Reeves, R.; Wolffe, A.P.
EMBO J. 15, 548-561, 1996
A;Title: Evidence for a shared structural role for HMGI and linker histones B4 and H1
A;Reference number: S62355; MUID:96174815; PMID:8599938
A;Accession: S62355
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-210 <NIG>
A;Cross-references: UNIPROT:Q91596; EMBL:U21933; NID:g709958; PIDN:AAC59859.1; PID:g709
A;Note: the authors did not translate the codon for residue 1
C;Superfamily: nonhistone chromosomal protein HMGI-2; HMGI box homology
F;8-82/Domain: HMGI box homology <HMGI>
F;91-165/Domain: HMGI box homology <HMGI>

Query Match 2.0%; Score 8; DB 2; Length 210;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 SDFRPKIK 293
|||||

```

Db 106 SDFRPKIK 113

RESULT 12

A99249

hypothetical protein fabG-1 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C:Accession: A99249

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweyez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: A99249

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-265 <KUR>

A:Cross-references: UNIPROT:Q97ZE4; GB:AE006641; NID:G13814158; PIDN:AAK41248.1; GSPDB:G

C:Genetics:

A:Gene: fabG-1

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 2.0%; Score 8; DB 2; Length 265;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 GKRVLVTA 201
|||||
Db 7 GKRVLVTA 14

RESULT 13

B97226

dihydroorotate dehydrogenase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C:Accession: B97226

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97226

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <KUR>

A:Cross-references: UNIPROT:Q97FS7; GB:AE001437; PIDN:AAK80597.1; PID:G15025678; GSPDB:G

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2650

C:Superfamily: dihydroorotate oxidase

Query Match 2.0%; Score 8; DB 2; Length 298;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 KELPKMKK 387
|||||
Db 82 KELPKMKK 89

RESULT 14

E71125

probable aromatic-amino-acid transaminase (EC 2.6.1.57) PH0771 [similarity] - Pyrococcus

C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004

C:Accession: E71125

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekir
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: E71125

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-391 <KAW>

A:Cross-references: UNIPROT:O58489; GB:AP000003; NID:G3236130; PIDN:BAA29863.1; PID:G323

A:Experimental source: strain OT3

A>Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH0771

C:Keywords: aminotransferase; phosphoprotein; pyridoxal phosphate

F:236/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 2.0%; Score 8; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TSKEKLEIE 342
|||||
Db 371 TSKEKLEIE 378

RESULT 15

T36548

probable transaminase (EC 2.6.1.-) SCH10.36 [similarity] - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: T36548

R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999

A:Reference number: Z21609

A:Accession: T36548

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-396 <MUR>

A:Cross-references: UNIPROT:Q9X8S5; EMEL:AL049754; PIDN:CAB42045.1; GSPDB:GN00070; SCOE

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB.SCH10.36

C:Superfamily: aspartate transaminase

C:Keywords: aminotransferase

Query Match 2.0%; Score 8; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVVTAGAT 204
|||||
Db 99 VLVVTAGAT 106

Search completed: July 2, 2005, 12:38:48
Job time : 24.3488 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:22:48 ; Search time 10.6047 Seconds
(without alignments)
1098.127 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKELIEPFSE.....PYRGNVQGSTRLAFSKRKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5E_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCUTS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	806	100.0	156	4	US-09-399-003-71
2	533	66.1	246	3	US-08-822-774-44
3	533	66.1	246	3	US-09-632-711-44
4	533	66.1	246	3	US-09-632-703B-44
5	533	66.1	246	4	US-09-632-702-44
6	533	66.1	246	4	US-09-399-003-44
7	185.5	23.0	197	4	US-09-328-352-4665
8	172	21.3	202	4	US-09-252-991A-17662
9	168.5	20.9	192	4	US-09-540-236-3681
10	168.5	20.9	204	4	US-09-543-681A-5063
11	164	20.3	217	4	US-09-489-039A-8299
12	160	19.9	42	3	US-08-822-774-41
13	160	19.9	42	3	US-09-632-711-41
14	160	19.9	42	3	US-09-632-703B-41
15	160	19.9	42	4	US-09-632-702-41
16	160	19.9	42	4	US-09-399-003-41
17	138	17.1	190	4	US-09-198-452A-412
18	138	17.1	190	4	US-09-438-185A-394
19	134.5	16.7	177	3	US-09-199-637A-247
20	121	15.0	147	4	US-09-583-110-4437
21	121	15.0	156	4	US-09-107-433-3527
22	119	14.8	27	3	US-08-822-774-38
23	119	14.8	27	3	US-09-632-711-38
24	119	14.8	27	3	US-09-632-703B-38
25	119	14.8	27	4	US-09-632-702-38
26	119	14.8	27	4	US-09-399-003-38
27	118	14.6	151	4	US-09-902-540-10618

28	114.5	14.2	138	2	US-08-824-405-11	Sequence 11, Appl
29	112	13.9	181	3	US-09-199-637A-239	Sequence 239, App
30	111	13.8	138	2	US-08-824-405-10	Sequence 10, Appl
31	111	13.8	164	2	US-08-824-405-2	Sequence 2, Appli
32	111	13.8	252	2	US-08-824-405-4	Sequence 4, Appli
33	111	13.8	252	4	US-09-538-092-1062	Sequence 1062, Ap
34	111	13.8	280	4	US-09-949-016-10934	Sequence 10934, A
35	102	12.7	178	3	US-09-171-461-2	Sequence 2, Appli
36	102	12.7	178	4	US-09-970-711-2	Sequence 2, Appli
37	99.5	12.3	254	3	US-09-199-637A-241	Sequence 241, App
38	87.5	10.9	183	3	US-09-134-001C-5336	Sequence 5336, Ap
39	87.5	10.9	1150	3	US-09-238-303-9	Sequence 9, Appli
40	87.5	10.9	1150	4	US-09-946-239-9	Sequence 9, Appli
41	87	10.8	147	2	US-08-824-405-12	Sequence 12, Appl
42	87	10.8	177	4	US-09-248-796A-17657	Sequence 17657, A
43	87	10.8	277	4	US-08-311-731A-284	Sequence 284, App
44	85.5	10.6	1031	3	US-08-811-682-15	Sequence 15, Appl
45	84	10.4	475	3	US-09-134-001C-4250	Sequence 4250, Ap

ALIGNMENTS

RESULT 1
US-09-399-003-71
; Sequence 71, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hogrefe, Holly
; APPLICANT: Hansen, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/ US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-399-003-71

Query Match 100.0%; Score 806; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.8e-93;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MLLPDWKIRKELIEPFSESLQAGYDLRVGREAFAVKGLIDVEKGVVIPPYVALI	60
Db	1	MLLPDWKIRKELIEPFSESLQAGYDLRVGREAFAVKGLIDVEKGVVIPPYVALI	60
Qy	61	LTLEIKLPDDVGMKIRSSLAEGVIGSFVWDPCWGNLTMLYNASNEPVELRYGE	120
Db	61	LTLEIKLPDDVGMKIRSSLAEGVIGSFVWDPCWGNLTMLYNASNEPVELRYGE	120
Qy	121	RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL	156
Db	121	RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL	156

RESULT 2
US-08-822-774-44
; Sequence 44, Application US/08822774
; Patent No. 618397
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)

```

; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; US-08-822-774-44

Query Match 66.1%; Score 533; DB 3; Length 246;
Best Local Similarity 98.1%; Pred. No. 2.5e-58;
Matches 102; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 53 PPREVALITLRIKLPDDVMGDMKIRSLAREGVIGSPAWVDPGWDGNLTMLYNASNE 112
DB 1 PPREVALITLRIKLPNNVMGDMKIRSLAREGVIGSPAWVDPGWDGNLTMLYNASNE 60

QY 113 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL 156
DB 61 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL 104

RESULT 3
US-09-632-711-44
; Sequence 44, Application US/09632711
; Patent No. 633165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000

```

```

; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-632-711-44

Query Match 66.1%; Score 533; DB 3; Length 246;
Best Local Similarity 98.1%; Pred. No. 2.5e-58;
Matches 102; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 53 PPREVALITLRIKLPDDVMGDMKIRSLAREGVIGSPAWVDPGWDGNLTMLYNASNE 112
DB 1 PPREVALITLRIKLPNNVMGDMKIRSLAREGVIGSPAWVDPGWDGNLTMLYNASNE 60

QY 113 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL 156
DB 61 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL 104

RESULT 4
US-09-632-703B-44
; Sequence 44, Application US/09632703B
; Patent No. 6379553
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF
; Protein Complexes, Isolated PEF Proteins, and Methods for
; Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: <Unknown>
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,703B
; FILING DATE: 24-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid

```



```
Db 35 WLDGRLNPRPPVERINGATVDVRLGNKFRTERGHTAPFDLSPKAEVSNALDRVMS 94
Qy 47 EGVKVIPT-----PREVALITLRIKLPDDVMGDMKIRSLAREGVIGSPA--WVDP 96
Db 95 E-EIVLPEGEAFPLPGELALAVTYESVTLPADLVGWLDRGSSLARLGLMVHTAHRIIDP 153
Qy 97 GWDGNLTMLYNASNEPVELRGERFVQIAFTIRLEGPARNPY 138
Db 154 GWSGCVILFYNKGLPLALRPGMIGALSFEPLSGPAARPY 195

RESULT 12
US-08-822-774-41
; Sequence 41, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-822-774-41

Query Match 19.9%; Score 160; DB 3; Length 42;
Best Local Similarity 96.9%; Pred. No. 1e-12;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVG 32
Db 11 MLLPDWKIRKEILIEPFSEWLQAGYDLRVG 42

RESULT 13
US-08-632-711-41
; Sequence 41, Application US/09632711
; Patent No. 6333165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
```

```
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSES: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-632-711-41

Query Match 19.9%; Score 160; DB 3; Length 42;
Best Local Similarity 96.9%; Pred. No. 1e-12;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVG 32
Db 11 MLLPDWKIRKEILIEPFSEWLQAGYDLRVG 42

RESULT 14
US-09-632-703B-41
; Sequence 41, Application US/09632703B
; Patent No. 6379553
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF
; Protein Complexes, Isolated PEF Proteins, and Methods for
; Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: <Unknown>
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,703B
; FILING DATE: 24-Aug-2000
```

```

;
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-632-703B-41

Query Match 19.9%; Score 160; DB 3; Length 42;
Best Local Similarity 96.9%; Pred. No. 1e-12;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVG 32
| | | | | | | | | | | | | | | | | | | |
Db 11 MLLPDWKIRKEILIEPFSEEWLQAGYDLRVG 42

Search completed: July 2, 2005, 12:26:00
Job time : 11.6047 secs

;
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-632-702-41

Query Match 19.9%; Score 160; DB 4; Length 42;
Best Local Similarity 96.9%; Pred. No. 1e-12;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVG 32
| | | | | | | | | | | | | | | | | | | |
Db 11 MLLPDWKIRKEILIEPFSEEWLQAGYDLRVG 42

Search completed: July 2, 2005, 12:26:00
Job time : 11.6047 secs

;
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-632-703B-41

Query Match 19.9%; Score 160; DB 3; Length 42;
Best Local Similarity 96.9%; Pred. No. 1e-12;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVG 32
| | | | | | | | | | | | | | | | | | | |
Db 11 MLLPDWKIRKEILIEPFSEEWLQAGYDLRVG 42

RESULT 15
US-09-632-702-41
; Sequence 41, Application US/09632702
; Patent No. 6444428
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,702
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:32:14 ; Search time 30.9767 Seconds
(without alignments)
2578.849 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 156
Sequence: 1 MLPLDWKIRKEILIEPFSEE.....PYRGNVQGSTRLAFSKRKKL 156

Scoring table:

Gapop 60.0 , Gapext 60.0
Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Uniprot 03.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	156	100.0	156	1 DCD_PYRFU	Q8x251 pyrococcus
2	156	100.0	156	2 Q6T180	Q6t180 pyrococcus
3	37	23.7	156	1 DCD_PYRHO	O57706 pyrococcus
4	22	14.1	154	1 DCD_PYRAB	Q9uxs8 pyrococcus
5	8	5.1	114	2 Q7V7Z6	Q7vvz6 bordetella
6	8	5.1	114	2 Q7W5U6	Q7w5u6 bordetella
7	8	5.1	114	2 Q7W5U7	Q7w5j7 bordetella
8	8	5.1	420	1 ODO2_BUCAI	P57389 buchnera ap
9	8	5.1	459	2 Q26001	Q26001 plasmidium
10	8	5.1	459	2 Q81501	Q81501 plasmidium
11	8	5.1	503	2 Q8UBB2	Q8ubb2 agrobacteri
12	7	4.5	34	2 Q9UQN9	Q9uqn9 homo sapien
13	7	4.5	37	2 Q9XQ05	Q9xqg5 toxoplasma
14	7	4.5	61	2 Q7P2B1	Q7p2b1 fusobacteri
15	7	4.5	64	1 SCX4_MESMA	P58328 mesobuthus
16	7	4.5	66	2 Q61ZE1	Q61zel mesobuthus
17	7	4.5	84	1 SC13_MESMA	Q9njc8 mesobuthus
18	7	4.5	84	1 SCXB_MESMA	Q9n882 mesobuthus
19	7	4.5	94	2 Q8XTK6	Q8xtk6 raistonias
20	7	4.5	127	2 Q8DUZ8	Q8duz8 streptococc
21	7	4.5	133	1 FLIS_BACSU	P39739 bacillus su
22	7	4.5	133	2 Q8ENI3	Q8eni3 oceanobacil
23	7	4.5	133	2 Q9K6W1	Q9k6w1 bacillus ha
24	7	4.5	137	2 Q6PRY2	Q6pry2 uncultured
25	7	4.5	144	2 Q9HFB4	Q9hfb4 rhizobium l
26	7	4.5	153	2 Q7W8B4	Q7w8b4 bordetella
27	7	4.5	153	2 Q7W1X2	Q7w1x2 bordetella
28	7	4.5	153	2 Q84CS3	Q84cs3 bordetella
29	7	4.5	160	2 Q8MRQ3	Q8mrq3 drosophila
30	7	4.5	163	2 Q6CU30	Q6cu30 kluyveromyc
31	7	4.5	164	2 Q8CR50	Q8cr50 staphylococ

32 7 4.5 175 2 Q6BIP7 O6bip7 debaryomyce
33 7 4.5 175 2 Q9UHT7 Q9unt7 homo sapien
34 7 4.5 177 2 Q8TH89 Q8th89 methanosarc
35 7 4.5 178 1 AIL_YEREN P16454 yersinia en
36 7 4.5 178 2 Q7OAM3 Q7oam3 yersinia en
37 7 4.5 180 2 Q6L1B1 Q6l1b1 picrophilus
38 7 4.5 182 1 AIL_YERPS O56957 yersinia ps
39 7 4.5 182 2 Q667X3 Q667x3 yersinia ps
40 7 4.5 182 2 Q74SP4 Q74sp4 yersinia pe
41 7 4.5 182 2 Q8ZCR3 Q8zcr3 yersinia pe
42 7 4.5 192 2 Q8EH22 Q8eh22 shewanella
43 7 4.5 194 2 Q8D0Z7 Q8d0z7 yersinia pe
44 7 4.5 198 2 Q9FMX7 Q9fmx7 arabidopsis
45 7 4.5 198 2 Q7NZY4 Q7nzy4 chromobacte

ALIGNMENTS

RESULT 1
ID DCD_PYRFU STANDARD; PRT; 156 AA.
AC Q8X251;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase).
GN Name=dcd; OrderedLocuNames=PF1996;
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21664347; PubMed-11782527; DOI=10.1073/pnas.012372799;
RA Hogrefe H.H., Hansen C.J., Scott B.R., Nielson K.B.;
RT "Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation."
RL Proc. Natl. Acad. Sci. U.S.A. 99:596-601(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate family.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY066005; AAL47572.1; -;
DR EMBL; AE010292; AAL82120.1; -;
DR HSSP; Q57872; LOGH.
DR HAMAP; MF_00146; -; 1.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 156 AA; 17869 MW; 00D57A5A027B451 CRC64;

Query Match 100.0%; Score 156; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e-154;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AJ248288; CAB50685.1; -;
 DR FIR; G75030; G75030.
 DR HSSP; Q57872; IOGH.
 DR HAMAP; MF 00146; -; 1.
 DR InterPro; IPR003232; dCTP deaminase.
 DR InterPro; IPR008180; DeoxyUTPase.
 DR Pfam; PF00692; dUTPase; 1.
 DR ProDom; PD004900; dCTP deaminase; 1.
 DR Complete proteome; Hydrolyase; Nucleotide metabolism.
 SQ SEQUENCE 154 AA; 17758 MW; C031BEE419094DDB CRC64;

Query Match 14.1%; Score 22; DB 1; Length 154;
 Best Local Similarity 100.0%; Pred. No. 2.1e-14; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 GSPAWDPGWDGNLTMLYNAS 110
 |||||
 DB 89 GSPAWDPGWDGNLTMLYNAS 110

RESULT 5
 Q7VYZ6 PRELIMINARY; PRT; 114 AA.
 AC Q7VYZ6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp1153;
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640414; CAE41450.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

Query Match 5.1%; Score 8; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 108 NASNEPVE 115
 |||||
 DB 57 NASNEPVE 64

RESULT 6
 Q7W5U6 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp3192;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640414; CAE41450.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

Query Match 5.1%; Score 8; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 NASNEPVE 115
 |||||
 DB 57 NASNEPVE 64

RESULT 6
 Q7W5U6 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp3192;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640414; CAE41450.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

ID Q7W5U6 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp3192;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=12822 / ATCC BAA-587;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640432; CAE38477.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

Query Match 5.1%; Score 8; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 NASNEPVE 115
 |||||
 DB 57 NASNEPVE 64

RESULT 7
 Q7W5U7 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U7;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative exported protein.
 GN OrderedLocusNames=BB3592;
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640447; CAE34086.1; -;

Query Match 5.1%; Score 8; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 108 NASNEPVE 115
 |||||
 DB 57 NASNEPVE 64

RESULT 6
 Q7W5U6 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp3192;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640447; CAE34086.1; -;

DR InterPro; IPR008972; Cupredoxin.
KW Complete proteome. 12335 MW; 731F7814735BA413 CRC64;
SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;
Query Match 5.1%; Score 8; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 NASNEPVE 115
Db 57 NASNEPVE 64
RESULT 8
ODO2_BUCAI STANDARD; PRT; 420 AA.
ID ODO2_BUCAI
AC P57389;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61) (E2)
DE (Dihydrolipoamide succinyltransferase component of 2-oxoglutarate dehydrogenase complex).
DE Name=sucB; OrderedLocusName=BU303;
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077; DOI=10.1038/35024074;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: The 2-oxoglutarate dehydrogenase complex catalyzes the overall conversion of 2-oxoglutarate to succinyl-CoA and CO(2). It contains multiple copies of three enzymatic components: 2-oxoglutarate dehydrogenase (E1), dihydrolipoamide succinyltransferase (E2) and lipoamide dehydrogenase (E3) (By similarity).
CC -!- CATALYTIC ACTIVITY: Succinyl-CoA + enzyme N(6) - (dihydrolipoal)lysine = CoA + enzyme N(6) - (S-succinylidihydrolipoal)lysine.
CC -!- COFACTOR: Contains 1 covalently bound lipoyl cofactor (By similarity).
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBUNIT: Forms a 24-polypeptide structural core with octahedral symmetry (By similarity).
CC -!- SIMILARITY: Belongs to the 2-oxoacid dehydrogenase family.
CC -!- SIMILARITY: Contains 1 lipoyl-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF001118; BAB13012.1; -.
CC HSP; P07016; 1C4T.
DR InterPro; IPR001078; 2Oxoacid dh.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR011053; Hybrid_motif.
DR InterPro; IPR003016; Lipoyl_BS.
DR InterPro; IPR006255; SucB.
DR Pfam; PF00198; 2-oxoacid dh; 1.
DR Pfam; PF00364; Biotin_lipoyl; 1.
DR ProDom; PD001115; 2Oxoacid_dh; 1.

DR TIGRFAMs; TIGR01347; sucB; 1.
DR PROSITE; PS00968; BIOTINYL_LIPOYL; 1.
DR PROSITE; PS00189; LIPOYL; FALSE_NEG.
KW Acyltransferase; Complete proteome; Lipoyl; Transferase;
KW Tricarboxylic acid cycle.
FT DOMAIN 1 77 Biotinyl/lipoyl.
FT BINDING 44 44 Lipoyl (potential).
FT ACT_SITE 391 391 Potential.
FT ACT_SITE 395 395 Potential.
SQ SEQUENCE 420 AA; 48091 MW; E0028D647A5CE34C CRC64;
Query Match 5.1%; Score 8; DB 1; Length 420;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 45 EKEGKVI 52
Db 62 EKEGKVI 69
RESULT 9
Q26001 PRELIMINARY; PRT; 459 AA.
ID Q26001
AC Q26001; (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE RabGDI protein.
DE Name=rabGDI;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=97001683; PubMed=8844675; DOI=10.1016/0166-6851(96)02606-0;
RA Attal G., Langsley G.;
RT "A Plasmodium falciparum homologue of a rab specific GDP dissociation inhibitor.";
RL Mol. Biochem. Parasitol. 79:91-95(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Langsley G. W.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X93166; CAA63653.1; -.
DR HSP; P21856; 1DST.
DR GO; GO:0005093; F:RAB GDP-dissociation inhibitor activity; IEA.
DR GO; GO:0015031; P:protein transport; IEA.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR000806; RabGDI.
DR InterPro; IPR002005; Rab_GDI_REP.
DR Pfam; PF00996; GDI; 1.
DR PRINTS; PR00892; RABGDI.
DR PRINTS; PR00891; RABGDIREP.
SQ SEQUENCE 459 AA; 52311 MW; DE1CCA47B124D8EA CRC64;
Query Match 5.1%; Score 8; DB 2; Length 459;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 61 LTLERIKL 68
Db 212 LTLERIKL 219
RESULT 10
Q81501 PRELIMINARY; PRT; 459 AA.
ID Q81501
AC Q81501; (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

AC Q9XQ05;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Ribosomal protein L36.
 OS Toxoplasma gondii.
 OG Chloroplast.
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
 OC Toxoplasma.
 OX NCBI_TaxID=5811;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kissinger J.C., Donald R.G., Moulton A.L., Gutell R., Aiello D.P.,
 RA Lang-Unaach N., Roos D.S.;
 RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to the L36P family of ribosomal proteins.
 DR EMBL; U87145; AAD1142.1; -.
 DR GO; GO:0009507; C:chloroplast; IEA.
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003735; P:protein biosynthesis; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR000473; Ribosomal_L36.
 DR Pfam; PF00444; Ribosomal_L36; 1.
 DR TIGRFAMs; TIGR01022; rpmJ_bact; 1.
 KW Chloroplast; Ribonucleoprotein; Ribosomal protein.
 SQ SEQUENCE 37 AA; 4546 MW; EDD8C59BEEB5E6AD CRC64;

Query Match 4.5%; Score 7; DB 2; Length 37;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 MKIRSSL 82
 |||||
 DB 1 MKIRSSL 7

RESULT 14
 Q7P2B1 PRELIMINARY; PRT; 61 AA.
 AC Q7P2B1;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN Name=FN0190;
 OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.
 OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
 OC Fusobacterium.
 OX NCBI_TaxID=209882;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 49256;
 RA Karpatal V., Ivanova N., Anderson I., Reznik G., Bhattacharyya A.,
 RA Gardner W.L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,
 RA Haselkorn R., Overbeek R., Kyrpides N.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 DR EMBL; AABF01000162; EAA23292.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 61 AA; 7335 MW; AF4936E8683E2348 CRC64;

Query Match 4.5%; Score 7; DB 2; Length 61;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 FSEESLQ 23
 |||||
 DB 9 FSEESLQ 15

RESULT 15

SCX4 MESMA
 ID SCX4 MESMA STANDARD; PRT; 64 AA.
 AC P58328;
 DT 28-PEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Alpha-like neurotoxin Bmk-IV (Bmk4) (Bmk M4) (BmkM4).
 OS Mesobuthus martensii (Machurian scorpion) (Buthus martensii).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
 OC Buthida; Buthoidea; Buthidae; Mesobuthus.
 OX NCBI_TaxID=34649;
 RN [1]
 RP SEQUENCE.
 RA TISSUE=Venom;
 RX MEDLINE=97346880; PubMed=9203297; DOI=10.1016/S0041-0101(96)00146-8;
 RA Luo M.-J., Xiong Y.-M., Wang M., Wang D.-C., Chi C.-W.;
 RT "Purification and sequence determination of a new neutral mammalian
 RT neurotoxin from the scorpion Buthus martensii Karsch.";
 RL Toxicon 35:723-731(1997).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.3 ANGSTROMS).
 RX MEDLINE=99425298; PubMed=10493862; DOI=10.1006/jmbi.1999.3036;
 RA He X.-L., Li H.-W., Zeng Z.-H., Liu X.-Q., Wang M., Wang D.-C.;
 RT "Crystal structures of two alpha-like scorpion toxins: non-proline cis
 RT peptide bonds and implications for new binding site selectivity on the
 RT sodium channel.";
 RL J. Mol. Biol. 292:125-135(1999).
 CC -!- FUNCTION: Binds to sodium channels and inhibits the inactivation
 CC of the activated channels, thereby blocking neuronal transmission.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
 CC -!- TOXIC DOSE: LD(50) is 4 mg/kg by intravenous injection into mice.
 CC -!- SIMILARITY: Belongs to the alpha/beta-scorpion toxin family.
 CC Alpha-toxin subfamily. Alpha-like toxin subclass.
 PDB; 1SN4; X-ray; A-
 DR InterPro; IPR003614; Kntot1.
 DR InterPro; IPR001219; Neurotoxin.
 DR InterPro; IPR002061; Scorpion_toxinL.
 DR Pfam; PF00537; Toxin_3; 1.
 DR PRINTS; PR00284; SCORPNTOXIN.
 DR PRINTS; PR00285; SCORPNTOXIN.
 DR ProDom; PD000908; Scorpion_toxinL; 1.
 DR SMART; SM00505; Kntot1; 1.
 KW 3D-structure; Direct protein sequencing; Ionic channel inhibitor;
 KW Neurotoxin; Sodium channel inhibitor; Toxin.
 FT DISULFID 12 63
 FT DISULFID 16 36
 FT DISULFID 22 46
 FT DISULFID 26 48
 FT STRAND 2 8
 FT TURN 9 11
 FT STRAND 12 13
 FT STRAND 15 15
 FT HELIX 19 28
 FT TURN 29 30
 FT STRAND 33 40
 FT TURN 41 42
 FT STRAND 43 51
 FT TURN 53 54
 FT STRAND 57 57
 SQ SEQUENCE 64 AA; 7021 MW; 20C7427C1C74DABB CRC64;

Query Match 4.5%; Score 7; DB 1; Length 64;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 IKLPDDV 72
 |||||
 DB 49 IKLPDDV 55

Search completed: July 2, 2005, 12:38:13
 Job time : 32.9767 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:22:48 ; Search time 27.3953 Seconds
(without alignments)
1098.127 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 1998
Sequence: 1 MLHHVKLIYATSKRLVGKK.....KMKRELAERIWDIEIKXLS 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1988	99.5	403	3 US-08-822-774-19	Sequence 19, Appl
2	1988	99.5	403	3 US-09-632-711-19	Sequence 19, Appl
3	1988	99.5	403	3 US-09-632-703B-19	Sequence 19, Appl
4	1988	99.5	403	4 US-09-632-702-19	Sequence 19, Appl
5	1988	99.5	403	4 US-09-399-003-19	Sequence 19, Appl
6	635.5	31.8	437	3 US-08-822-774-51	Sequence 51, Appl
7	635.5	31.8	437	3 US-09-632-711-51	Sequence 51, Appl
8	635.5	31.8	437	3 US-09-632-703B-51	Sequence 51, Appl
9	635.5	31.8	437	4 US-09-632-702-51	Sequence 51, Appl
10	635.5	31.8	437	4 US-09-399-003-51	Sequence 51, Appl
11	466	24.3	408	4 US-09-252-991A-24057	Sequence 24057, A
12	466	24.3	418	4 US-09-543-681A-7435	Sequence 7435, Ap
13	461.5	23.1	399	4 US-09-710-279-2576	Sequence 2576, Ap
14	461.5	23.1	400	3 US-09-134-001C-4785	Sequence 4785, Ap
15	458.5	22.9	431	4 US-09-540-236-3536	Sequence 3536, Ap
16	439	22.0	427	4 US-09-328-352-5205	Sequence 5205, Ap
17	438.5	21.9	444	3 US-08-822-774-52	Sequence 52, Appl
18	438.5	21.9	444	3 US-09-632-711-52	Sequence 52, Appl
19	438.5	21.9	444	3 US-09-632-703B-52	Sequence 52, Appl
20	438.5	21.9	444	4 US-09-632-702-52	Sequence 52, Appl
21	438.5	21.9	444	4 US-09-399-003-52	Sequence 52, Appl
22	418.5	20.9	398	4 US-09-489-039A-10656	Sequence 10656, A
23	312.5	13.6	280	4 US-09-902-540-12020	Sequence 12020, A
24	262	13.1	183	4 US-09-583-110-4578	Sequence 4578, Ap
25	236	11.8	252	4 US-09-583-110-4579	Sequence 4579, Ap
26	235.5	11.8	245	4 US-09-107-532A-4448	Sequence 4448, Ap
27	235	11.8	188	4 US-09-107-532A-6818	Sequence 6818, Ap

28	233	11.7	262	4 US-09-134-000C-3905	Sequence 3905, Ap
29	219	11.0	154	3 US-09-107-433-4272	Sequence 4272, Ap
30	158	7.9	188	3 US-09-627-376-9	Sequence 9, Appl
31	158	7.9	188	4 US-10-047-676B-9	Sequence 9, Appl
32	151.5	7.6	181	2 US-08-645-193B-24	Sequence 24, Appl
33	149.5	7.5	181	2 US-08-392-625-22	Sequence 22, Appl
34	149.5	7.5	181	2 US-08-466-961A-22	Sequence 22, Appl
35	139	7.0	358	4 US-09-248-796A-15014	Sequence 15014, A
36	135	6.8	31	3 US-08-822-774-46	Sequence 46, Appl
37	135	6.8	31	3 US-09-632-711-46	Sequence 46, Appl
38	135	6.8	31	3 US-09-632-703B-46	Sequence 46, Appl
39	135	6.8	31	4 US-09-632-703-46	Sequence 46, Appl
40	135	6.8	31	4 US-09-399-003-46	Sequence 46, Appl
41	135	6.8	459	4 US-09-248-796A-19036	Sequence 19036, A
42	134	6.7	1500	3 US-09-323-472A-2	Sequence 2, Appl
43	134	6.7	1500	3 US-09-323-472A-4	Sequence 4, Appl
44	134	6.7	1500	3 US-09-323-472A-12	Sequence 12, Appl
45	134	6.7	1500	3 US-09-323-472A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-822-774-19
; Sequence 19, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREPE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSES: David J. Kulik, Evenson, McKeown, Edwards &
; Lanehan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-822-774-19

Query Match 99.5%; Score 1988; DB 3; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4e-181;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLHHVKLIYATSKRLVGKKIVXXXPGSIAALDVKACEGLRHGAHVHVMSEAATKIIH 60
Db 1 MLHHVKLIYATSKRLVGKKIVXXXPGSIAALDVKACEGLRHGAHVHVMSEAATKIIH 60

```
QY 61 PYAWNLPCTGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIACGIDDPVTT 120
DB 61 PYAWNLPCTGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIACGIDDPVTT 120
QY 121 VTTTAPPHIPIMAPAMHETMYRHPVIRENIERLKLGVFEFIGPRIEGRKAVASIDEIV 180
DB 121 VTTTAPPHIPIMAPAMHETMYRHPVIRENIERLKLGVFEFIGPRIEGRKAVASIDEIV 180
QY 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFITNASSGKMGVALAEADPRGAVTLIR 240
DB 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFITNASSGKMGVALAEADPRGAVTLIR 240
QY 241 TKGSKVAFRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKASGKIKS 300
DB 241 TKGSKVAFRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKASGKIKS 300
QY 301 GRSITIELVPXNPKIIDRIKETQPNVFLVGPKAETSKKLIBEGKQRIERAKADLVVNT 360
DB 301 GRSITIELVPXNPKIIDRIKETQPNVFLVGPKAETSKKLIBEGKQRIERAKADLVVNT 360
QY 361 LEAFGSEENQVVLIGRDTFKLPKMKKRELAERIWDEIEKKLS 403
DB 361 LEAFGSEENQVVLIGRDTFKLPKMKKRELAERIWDEIEKKLS 403

RESULT 2
US-09-632-711-19
; Sequence 19, Application US/09632711
; Patent No. 633165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
;
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSER: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-632-711-19
Query Match 99.5%; Score 1988; DB 3; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4e-181;
```

```
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHHVKLLIYATKSRKLVGKKIVXXXPGSIALDVKACEGLIRHGAEVHVMSEAAATKI 60
DB 1 MLHHVKLLIYATKSRKLVGKKIVXXXPGSIALDVKACEGLIRHGAEVHVMSEAAATKI 60
QY 61 PYAWNLPCTGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIACGIDDPVTT 120
DB 61 PYAWNLPCTGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIACGIDDPVTT 120
QY 121 VTTTAPPHIPIMAPAMHETMYRHPVIRENIERLKLGVFEFIGPRIEGRKAVASIDEIV 180
DB 121 VTTTAPPHIPIMAPAMHETMYRHPVIRENIERLKLGVFEFIGPRIEGRKAVASIDEIV 180
QY 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFITNASSGKMGVALAEADPRGAVTLIR 240
DB 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFITNASSGKMGVALAEADPRGAVTLIR 240
QY 241 TKGSKVAFRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKASGKIKS 300
DB 241 TKGSKVAFRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKASGKIKS 300
QY 301 GRSITIELVPXNPKIIDRIKETQPNVFLVGPKAETSKKLIBEGKQRIERAKADLVVNT 360
DB 301 GRSITIELVPXNPKIIDRIKETQPNVFLVGPKAETSKKLIBEGKQRIERAKADLVVNT 360
QY 361 LEAFGSEENQVVLIGRDTFKLPKMKKRELAERIWDEIEKKLS 403
DB 361 LEAFGSEENQVVLIGRDTFKLPKMKKRELAERIWDEIEKKLS 403

RESULT 3
US-09-632-703B-19
; Sequence 19, Application US/09632703B
; Patent No. 6379553
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF
; Protein Complexes, Isolated PEF Proteins, and Methods for
; Identifying Same
;
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,703B
; FILING DATE: 24-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
```



```

;
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-632-703B-19
Query Match 99.5%; Score 1988; DB 3; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4e-181;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHHVKLIIVATSKRKLGVKKIVXXXPSTAAALDVKACEGLIRHGAHVHVMSEAATKIIH 60
DB 1 MLHHVKLIIVATSKRKLGVKKIVXXXPSTAAALDVKACEGLIRHGAHVHVMSEAATKIIH 60
QY 61 PYAWNLPTGNPVITETITGFIHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120
DB 61 PYAWNLPTGNPVITETITGFIHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120
QY 121 VVTTAPPHIPIMIAPAMHETMYRHPVIRENIERLKKLGVFEFGPRIEGRKVASIDEIV 180
DB 121 VVTTAPPHIPIMIAPAMHETMYRHPVIRENIERLKKLGVFEFGPRIEGRKVASIDEIV 180
QY 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFTNASSGKMGVALAEADPRGAVTLIR 240
DB 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFTNASSGKMGVALAEADPRGAVTLIR 240
QY 241 TGSKVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKIKS 300
DB 241 TGSKVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKIKS 300
QY 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360
DB 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360
QY 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELABRIWDEIEKXLS 403
DB 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELABRIWDEIEKXLS 403
RESULT 4
US-09-632-702-19
; Sequence 19, Application US/09632702
; Patent No. 6444428
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.B.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,702
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 19:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-632-702-19
Query Match 99.5%; Score 1988; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4e-181;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHHVKLIIVATSKRKLGVKKIVXXXPSTAAALDVKACEGLIRHGAHVHVMSEAATKIIH 60
DB 1 MLHHVKLIIVATSKRKLGVKKIVXXXPSTAAALDVKACEGLIRHGAHVHVMSEAATKIIH 60
QY 61 PYAWNLPTGNPVITETITGFIHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120
DB 61 PYAWNLPTGNPVITETITGFIHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120
QY 121 VVTTAPPHIPIMIAPAMHETMYRHPVIRENIERLKKLGVFEFGPRIEGRKVASIDEIV 180
DB 121 VVTTAPPHIPIMIAPAMHETMYRHPVIRENIERLKKLGVFEFGPRIEGRKVASIDEIV 180
QY 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFTNASSGKMGVALAEADPRGAVTLIR 240
DB 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFTNASSGKMGVALAEADPRGAVTLIR 240
QY 241 TGSKVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKIKS 300
DB 241 TGSKVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKIKS 300
QY 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360
DB 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360
QY 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELABRIWDEIEKXLS 403
DB 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELABRIWDEIEKXLS 403
RESULT 5
US-09-399-003-19
; Sequence 19, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hogsafe, Holly
; APPLICANT: Hansen, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; and Methods for Purifying and Identifying Same
; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying Same
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(403)
; OTHER INFORMATION: "x" represents any amino acid
US-09-399-003-19
Query Match 99.5%; Score 1988; DB 4; Length 403;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-632-702-19
Query Match 99.5%; Score 1988; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4e-181;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHHVKLIIVATSKRKLGVKKIVXXXPSTAAALDVKACEGLIRHGAHVHVMSEAATKIIH 60
DB 1 MLHHVKLIIVATSKRKLGVKKIVXXXPSTAAALDVKACEGLIRHGAHVHVMSEAATKIIH 60
QY 61 PYAWNLPTGNPVITETITGFIHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120
DB 61 PYAWNLPTGNPVITETITGFIHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120
QY 121 VVTTAPPHIPIMIAPAMHETMYRHPVIRENIERLKKLGVFEFGPRIEGRKVASIDEIV 180
DB 121 VVTTAPPHIPIMIAPAMHETMYRHPVIRENIERLKKLGVFEFGPRIEGRKVASIDEIV 180
QY 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFTNASSGKMGVALAEADPRGAVTLIR 240
DB 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFTNASSGKMGVALAEADPRGAVTLIR 240
QY 241 TGSKVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKIKS 300
DB 241 TGSKVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKIKS 300
QY 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360
DB 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360
QY 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELABRIWDEIEKXLS 403
DB 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELABRIWDEIEKXLS 403
```


Extracts, PEF Protein Complexes, Isolated PEF Proteins,
and Methods for Purifying and Identifying Same

NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,702
FILING DATE: 04-Aug-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997

ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844

INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 437 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Methanococcus Jannaschii
SEQUENCE DESCRIPTION: SEQ ID NO: 51:

US-09-632-702-51

Query Match 31.8%; Score 635.5; DB 4; Length 437;
Best Local Similarity 37.9%; Pred. No. 6e-52;
Matches 162; Conservative 83; Mismatches 153; Indels 29; Gaps 12;

QY 1 MLHVKLIYATKSRKLVGKKIVXXXPGSIAALDV-KACEGLIRHGAEVHVNSEAATKII 59
DB 5 IMHPTKLLKGTSKLLNKKILVAVTSSIAAETPKLMRELIRHGAEVVCIITEETKII 64

QY 60 HPYANLPTGNPVITEITGF-----IEHVELAGEHENKADLLVCPATANTISKIACGID 114
DB 65 GKEALKFGCGNEVYEITGXXXXDIEHLLYXXXXNECDCLLIYPATANIISKINLGA 124

QY 115 DTPVTVTVTAPPHIPIMAPAMHETMYRHPIVRENIERLK-KLGVEFTGPRIEE---- 168
DB 125 DNIVNTTALMPFGNKPFIIVPAMHENMFXXAIRKHKIDKLEKDKIYIISPKFEEXXXX 184

QY 169 -GRAKVASIDEIVRVVIKKL-HKKTLEGRVLTAGATREYIDPIRFTNASSGKMGVAL 226
DB 185 XGKAIVANIEDVVKAVIEKIGNLKKEGNRVLILNGGTVEFDKVRVISNLSGKMGVAL 244

QY 227 ABEADPRG-AVTLIRTKGSVKAFRIRKIKLVETVEEMLS-AIENELRSKDYDVVIMAAA 284
DB 245 ABAFCCKGFEYVEVITAMGLEPPYIK--NHKVLTAKEMLNKAIEXXLAKDPDIISSAA 302

QY 285 VSDPRPKIAEGKIKS----GRSITIELVPXNPKIIDRIKETQPNV-FLVGKAE--TSK 337
DB 303 ISDFTVESKFEGLKSEEEEXXXLLKLRXNPKVLELRRIYKDKVIIGKAEYNLDE 362

QY 338 EKLIEGKRQIRAKADLVVGNLTLEA----FGSEENQVVLIGRDTFKELPKMKKRELAE 393
DB 363 KELINRAKERLNKYNLNMIIANDLSKXXHYFGDDVIEVVIITKYEVEKISGSKKXKIEISER 422

QY 394 IWDEIEK 400
DB 423 IVEKVKK 429

RESULT 10
US-09-399-003-51
Sequence 51, Application US/09399003
Patent No. 6734293
GENERAL INFORMATION:
APPLICANT: Hansen, Connie J
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
FILE REFERENCE: 4121.0116-02
CURRENT APPLICATION NUMBER: US/09/399,003
CURRENT FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: PCT/ US98/05497
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: US 08/957,709
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: US 08/822,774
PRIOR FILING DATE: 1997-03-21
NUMBER OF SEQ ID NOS: 89
SOFTWARE: PatentIn version 3.1
SEQ ID NO 51
LENGTH: 437
TYPE: PRT
ORGANISM: Pyrococcus furiosus
NAME/KEY: MISC FEATURE
LOCATION: (1)-(437)
OTHER INFORMATION: "X" represents any amino acid
US-09-399-003-51

Query Match 31.8%; Score 635.5; DB 4; Length 437;
Best Local Similarity 37.9%; Pred. No. 6e-52;
Matches 162; Conservative 83; Mismatches 153; Indels 29; Gaps 12;

QY 1 MLHVKLIYATKSRKLVGKKIVXXXPGSIAALDV-KACEGLIRHGAEVHVNSEAATKII 59
DB 5 IMHPTKLLKGTSKLLNKKILVAVTSSIAAETPKLMRELIRHGAEVVCIITEETKII 64

QY 60 HPYANLPTGNPVITEITGF-----IEHVELAGEHENKADLLVCPATANTISKIACGID 114
DB 65 GKEALKFGCGNEVYEITGXXXXDIEHLLYXXXXNECDCLLIYPATANIISKINLGA 124

QY 115 DTPVTVTVTAPPHIPIMAPAMHETMYRHPIVRENIERLK-KLGVEFTGPRIEE---- 168
DB 125 DNIVNTTALMPFGNKPFIIVPAMHENMFXXAIRKHKIDKLEKDKIYIISPKFEEXXXX 184

QY 169 -GRAKVASIDEIVRVVIKKL-HKKTLEGRVLTAGATREYIDPIRFTNASSGKMGVAL 226
DB 185 XGKAIVANIEDVVKAVIEKIGNLKKEGNRVLILNGGTVEFDKVRVISNLSGKMGVAL 244

QY 227 ABEADPRG-AVTLIRTKGSVKAFRIRKIKLVETVEEMLS-AIENELRSKDYDVVIMAAA 284
DB 245 ABAFCCKGFEYVEVITAMGLEPPYIK--NHKVLTAKEMLNKAIEXXLAKDPDIISSAA 302

QY 285 VSDPRPKIAEGKIKS----GRSITIELVPXNPKIIDRIKETQPNV-FLVGKAE--TSK 337
DB 303 ISDFTVESKFEGLKSEEEEXXXLLKLRXNPKVLELRRIYKDKVIIGKAEYNLDE 362

QY 338 EKLIEGKRQIRAKADLVVGNLTLEA----FGSEENQVVLIGRDTFKELPKMKKRELAE 393
DB 363 KELINRAKERLNKYNLNMIIANDLSKXXHYFGDDVIEVVIITKYEVEKISGSKKXKIEISER 422

QY 394 IWDEIEK 400
DB 423 IVEKVKK 429

RESULT 11

US-09-252-991A-24057
; Sequence 24057, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24057
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24057

Query Match 24.5%; Score 489; DB 4; Length 405;
Best Local Similarity 34.8%; Pred. No. 4.7e-38;
Matches 136; Conservative 77; Mismatches 132; Indels 46; Gaps 16;

Qy 14 RKLVGKIVXXPGSIAALDVKACEGLIR----HGAEVHVMSEAATKIIHPYAMNLTG 69
Db 5 QRLYRKRIVLGVGGGIAA--YKSAE-LVRLRDQGAEEVRVMTQGGREFITPLTLQALSG 61
Qy 70 NPVITEI-----TGFIEHVELAGEHENKADLLVCPATANTISKIACGDDTPTVTVT 124
Db 62 HPVHTDLLDPAEAAAMGHIELA----RWADLVIAPATADLMARLVQGVANDLLTTLVLA 117
Qy 125 AFPHIPMIAPAMHETMYRHPVIRENIERLKLGVFEIGPRI-----EGRKAVASIDE 178
Db 118 T--DAQIALAPAMQAMWRDTATQANELLRQGFHLFGPAAGSQAGDVGGLRMLEAE 175
Qy 179 IVRVIKLHKHTLEGRVLVTAGATREYIDIRFITNASSGKMGVALAEADFRGA-VT 237
Db 176 LAQRAADCQFQALTGTVHLITAGTQENIDPVRYITNHSKGWGFALAEAAVEAGARVT 235
Qy 238 LIRTKSV---KAFRTRKIKLVETVEEMLSAIENELRSKKYDVVIMAAAVSDFRPKIKA 294
Db 236 LV--TGPVHLPTDRQVRV--DVVSARDMLAACEAEM---PCDLLIASAAVADYRPEVA 288
Qy 295 EGKIK---SGRSITTELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEGKQIER 350
Db 289 AHLKKDPTSGEGLLIQLV-RNPDILATLAQRDRPFSVGFAAET--ENLLDYAARKIKD 345
Qy 351 AKADLVVGNLT-----AFGSEENVVVLIGRD 377
Db 346 KNLDLVANDVANPSIGFNSDENAITVIDRD 376

RESULT 12

US-09-543-681A-7435
; Sequence 7435, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7435
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Proteus mirabilis

US-09-543-681A-7435

Query Match 23.3%; Score 466; DB 4; Length 418;
Best Local Similarity 34.2%; Pred. No. 7.7e-36;
Matches 149; Conservative 76; Mismatches 145; Indels 66; Gaps 20;
Qy 5 VKL---IYATK---SRKLVGKIVXXPGSIAALDVKACEGLIRH-----CAEVHVMSEA 54
Db 3 VKLWTFIYAIRHFIMTTLHDKNIIIGISGIAA--YKRAPE-LVRLRDKGAIVRVWMTPA 59
Qy 55 ATKIIHPYAMNLTGNPVITEI-----TGFIEHVELAGEHENKADLLVCPATANTISKI 109
Db 60 AHAFVTPLSIQAVSGFPVADLLDPAEAAAMGHIELG----KWADLILLAPATADLIARL 115
Qy 110 ACGIDTPTVTVVTTAPPHIPIMAPAMHETMYRHPVIRENIERLKLGVFEIGP----- 164
Db 116 RMGMANDLLTTLCLAS--AAPAIAPAMNQMYRATITQENLTALAQRCGLIWGDSGSQ 173
Qy 165 ---RIEGR---AKVASIDEIVRVIKLHKHTLEGRVLVTAGATREYIDIRFITNA 217
Db 174 ACGDVGPGRMLDPLVLAEOQFAI-----QHDFGKKITITAGPTREALDPVRFISNH 228
Qy 218 SSGKMGVALAEADFRGA-VTLIR-----TKGSVKAFRIKIKLVETVEEMLSAIENE 270
Db 229 SSGKMGFATAQAARQGAEEVTLIAGPVTLPTPACVK-----RIDVESAEQMYQQWMDI 281
Qy 271 LRSKYDVVIMAAVSDFRPKIKAEGKI-KSGRSITTELVPXNPKIIDRI-KBIQPNVFL 328
Db 282 AQSQ--DIFIGCAAVADYRAKQVAPEKIKQGDDEVITMI-KNPDIVASVGKMHVHRPFV 338
Qy 329 VGFKATSEKLEIEGKQIERAKADLVVGNLT-----AFGSEENVVVLIGRFTKELPK 384
Db 339 VGFAAET--QNVVEYARRKEQKQLDICANDVSLKDGAFNSDNALHLIANGETRLPH 396
Qy 385 MKKRELAERIWDIEIK 400
Db 397 SKTQLSHRLDDEIAK 412

RESULT 13

US-09-710-279-2576
; Sequence 2576, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2576
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-2576

Query Match 23.1%; Score 461.5; DB 4; Length 399;

Best Local Similarity 34.3%; Pred. No. 1.9e-35;
Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;
Qy 19 KKIIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKIIHPYAMNLTGNPVIT--- 74
Db 2 KHLIAVTGGIAAYKAIDLTSKLIQSGYDVRVWLSDBAQEFVFTPLAFQAISRPNVYNTF 61
Qy 75 --EITGFIEHVELAGEHENKADLLVCPATANTISKIACGI-DDTPTVTVTVTAFPHIPI 131
Db 62 KEENPEEIQHVSLG-----DWADAIIVAPATANTIAKLSVGIADDLITSTLLATTPK--- 114

QY 132 MIAPAMHETMYRHPVIRENIEIRKLVGEFICP-----RIEGRKVASIDE 178
Db 115 FVAPAMVNNYNNPRTKHKMVLSDQGYFFIEPGSGYLACGVAKGRMEE-PMQILSVIN 173
QY 179 IYVRVKKLHKHTLEGRKVLVTAGATREVIDPRTITNASSGKMGVALAEADFRGA-VT 237
Db 174 KEFTQKNNVKSFSFGKRALVTAGTVEVIDPVRVYNNRSGKMGYATAEARLDRGAIVT 233
QY 238 LIKTSGSVKAFIRKIK-LKVETVEEMLSAIENELRSKKYDVVIMAAVSDRPRPKIKABG 296
Db 234 LI--SGPHTLSLPEGINNKKVSVESADDMFOAVTE--RFAKQDIVIKAAVSDVTPMDILEH 289
QY 297 KIK---SGRSITIELVPXNPKIIDRIKEIOPNVFLGVGFAETSKELIEEGKQIERAKA 353
Db 290 KKKQEGGLSVQFK---RTKDILKYLGENKTHOYLVGFAAET--QNIQYALDKLKRKNA 344
QY 354 DLV----VGNLTLEAFSGSEENQVVLIGRDTFK-ELPKMKKRELAERIWDEIE 399
Db 345 DVIISNNVGDTSIGFSSDDNLTMHFKNNEKVNKKKKSALAHQIIELE 395

RESULT 14
US-09-134-001C-4785 Query Match 23.1%; Score 461.5; DB 3; Length 400;
; Sequence 4785, Application US/09134001C Best Local Similarity 34.3%; Pred. No. 1.9e-35;
; Patent No. 6380370 Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007 EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4785
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4785

QY 19 KIVVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAATKIHPYAWNLPNGPVIT--- 74
Db 3 KHLIAVTGGIAAYKAIDUTSKLIQSGYDVRVWLSHQAQEFVTPLAFOAISRNPTVTNTF 62
QY 75 --EITGFIHVELAGHENKADLILVCPATANTISKIACGI-DDTPVTVTVTATPPHIPI 131
Db 63 KEENPEIQHVSIG---DWADAIIVAPATANTIAKLSVGIADLLITSLTATTPK--- 115
QY 132 MIAPAMHETMYRHPVIRENIEIRKLVGEFICP-----RIEGRKVASIDE 178
Db 116 FVAPAMVNNYNNPRTKHKMVLSDQGYFFIEPGSGYLACGVAKGRMEE-PMQILSVIN 174
QY 179 IYVRVKKLHKHTLEGRKVLVTAGATREVIDPRTITNASSGKMGVALAEADFRGA-VT 237
Db 175 KEFTQKNNVKSFSFGKRALVTAGTVEVIDPVRVYNNRSGKMGYATAEARLDRGAIVT 234
QY 238 LIKTSGSVKAFIRKIK-LKVETVEEMLSAIENELRSKKYDVVIMAAVSDRPRPKIKABG 296
Db 235 LI--SGPHTLSLPEGINNKKVSVESADDMFOAVTE--RFAKQDIVIKAAVSDVTPMDILEH 290
QY 297 KIK---SGRSITIELVPXNPKIIDRIKEIOPNVFLGVGFAETSKELIEEGKQIERAKA 353
Db 291 KKKQEGGLSVQFK---RTKDILKYLGENKTHOYLVGFAAET--QNIQYALDKLKRKNA 345
QY 354 DLV----VGNLTLEAFSGSEENQVVLIGRDTFK-ELPKMKKRELAERIWDEIE 399

Db 346 DVIISNNVGDTSIGFSSDDNLTMHFKNNEKVNKKKKSALAHQIIELE 396
RESULT 15
US-09-540-236-3536 Query Match 22.9%; Score 458.5; DB 4; Length 431;
; Sequence 3536, Application US/09540236 Best Local Similarity 33.0%; Pred. No. 4.2e-35;
; Patent No. 6673910 Matches 140; Conservative 88; Mismatches 147; Indels 49; Gaps 19;
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001 FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3536
; LENGTH: 431
; TYPE: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-3536

QY 5 VKLIYATKSRKLVGK--KIVVXXPGSIAAL-DVKACEGLIRHGAHVHVMSEAATKIHP 61
Db 11 IMMFAKCOMKVMTOQLPKILLAVTGGIAAYKSAMLARLLIKSGCQVRVMTAGACEFITP 70
QY 62 YAWNLPNTGNPVITEITG-----FIEHVELAGHENKADLILVCPATANTISKIACGIDDT 116
Db 71 LTFQALTCQAVHTKLLDDEAERGGMGHITLA---KWADLVVVA PASANTIAKLAGGFADN 126
QY 117 PVTVTVTAFPHIPIMAPAMHETMYRHPVIRENIEIRKLVGEFICPRIEE-----GR 170
Db 127 LVTTVCLAT--DAPILAPAMNQMWANAIVQDNLKLKLRQFGYHIMTPDSGQACGDVGL 184
QY 171 AKVASIDEIVYRVI---KKLH-KKTLGKRVLTAGATREYIDPRTITNASSGKMGVAL 226
Db 185 ORLPEPEDLCOQILAFDKCRHIQPSLLGKTVVITAGATVEPIDPVRFLSNHSTGKMGYAL 244
QY 227 AEEADFRGA-VTLIRTKG-SVKA-FRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAA 283
Db 245 ANACYHAGADYLVSGKHVSLKTPFGVR--RLDVGTAEQMLKVCLSV--CQKADIFIAAA 300
QY 284 AVSDFRPKIKABGKI---KSGRSITIELVPXNPKIIDRIKEIOPNVFLGVGFAET----- 335
Db 301 AVADFKAAASVADHKIKTKTKNQOTMTLQLI-KNPDLVLTATICTHTYDPLICVGFAAETQDANN 359
QY 336 -SKEKLIIEGKQIERAKADLVVGNLTLEAFSGSEENQVVLIGRD---FTKELPKMKKREL 390
Db 360 CAKSKL---AAKQLDMIADVNSDKTI-GFGSENNAMTVFFAEQYDQMPQNLPKAHKSKI 415
QY 391 AERI 394
Db 416 AEQL 419

Search completed: July 2, 2005, 12:25:59
Job time : 29.3953 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 11:27:13 ; Search time 31.2558 Seconds
(without alignments)
2555.824 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKELIEPFSEE.....PYRGNVQSGSTRLAFSKRKKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	806	100.0	156	1 DCD_PYRFU	Q8x251 pyrococcus
2	806	100.0	156	2 Q6T180	Q6t180 pyrococcus
3	768	95.3	156	1 DCD_PYRAB	O57706 pyrococcus
4	703	87.2	154	1 DCD_PYRAB	Q9uxs8 pyrococcus
5	257.5	31.9	166	1 DCD_TTEAC	Q9hkk0 thermoplasma
6	245	30.4	172	2 Q684E1	Q684e1 sulfolobus
7	244	30.3	158	2 Q8V9P5	Q8v9p5 sulfolobus
8	243	30.1	180	1 DCD_AQUAE	O67539 aquifex ae
9	242.5	30.1	177	1 DCD_BACHD	Q9kfv3 bacillus ha
10	242	30.0	173	1 DCD_ACIAM	Q02103 acidianus a
11	237	29.4	158	2 Q71028	O71028 archaeal vi
12	237	29.4	158	2 Q777M3	O777m3 sulfolobus
13	236	29.3	176	1 DCD_PYRAE	Q8zw23 pyrobaculum
14	233	28.9	181	2 Q72G14	Q72g14 thermus the
15	233	28.9	183	1 DCD_SULTO	Q976g3 sulfolobus
16	232.5	28.8	181	1 DCD_ASRPE	Q9vfa8 aeropyrum p
17	232	28.8	143	2 Q612T7	Q612t7 picophilus
18	227.5	28.2	198	2 Q74MA7	Q74ma7 nanocarchaeu
19	223.5	27.7	157	2 Q97B24	Q97b24 thermoplasma
20	222.5	27.6	190	2 Q73T98	Q73t98 methanopyru
21	220	27.3	193	1 DCD_METKA	Q8tky5 methanopyru
22	218	27.0	189	1 DCD_CORGL	Q8nlt9 corynebacte
23	216	26.8	191	1 DCD_STRCO	Q9x8w0 streptomyc
24	214	26.6	191	1 DCD_STRAW	Q8a2z9 streptomyc
25	208	25.8	187	2 Q6NEW7	Q6new7 corynebacte
26	207.5	25.7	184	1 DCD_SULSO	Q980t7 sulfolobus
27	206	25.6	193	1 DCD_COREF	Q8fm44 corynebacte
28	205.5	25.5	196	1 DCD_TROW8	Q83h71 tropheryma
29	205.5	25.5	196	1 DCD_TROWT	Q820x8 tropheryma
30	203.5	25.2	201	2 Q6AC71	Q6ac71 leifsonia x
31	199	24.7	197	1 DCD_METTH	O27875 methanobact

32	198	24.6	190	1 DCD_MYCBO	Q7u297 mycobacteri
33	198	24.6	190	1 DCD_MYCTU	O07247 mycobacteri
34	197.5	24.5	190	1 DCD_MYCLE	Q9cb17 mycobacteri
35	191.5	23.8	193	1 DCD_BIFLO	Q8g478 bifidobacte
36	189.5	23.5	186	1 DCD_CAMJE	Q9pn07 campylobact
37	185	23.0	193	1 DCD_BUCBP	P59464 buchnera ap
38	184.5	22.9	195	1 DCD_HALN1	Q9hsq3 halobacteri
39	181	22.5	188	1 DCD_HELPY	O25136 helicobacte
40	180	22.3	188	1 DCD_HELPU	Q9zkd0 helicobacte
41	179.5	22.3	188	1 DCD_HELHP	Q7vj1q helicobacte
42	177	22.0	193	1 DCD_YERPE	Q8zfz8 yersinia pe
43	177	22.0	193	2 Q66C69	Q66c69 yersinia ps
44	174	21.6	193	1 DCD_BUCAI	P57209 buchnera ap
45	172.5	21.4	189	2 Q6FE29	Q6fe29 acinetobact

ALIGNMENTS

RESULT 1

ID DCD_PYRFU STANDARD; PRT; 156 AA.
AC Q8X251;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
DE deaminase).
GN Name=dcD; OrderedLocusNames=PF1996;
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21664347; PubMed=11782527; DOI=10.1073/pnas.012372799;
RA Hogrefe H.H., Hansen C.J., Scott B.R., Nielson K.B.;
RT "Archaeal dUTPase enhances PCR amplifications with archaeal DNA
RT polymerases by preventing dUTP incorporation.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:596-601(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY056605; AAL47572.1; -;
DR EMBL; AE010292; AAL82120.1; -;
DR HSSP; Q57872; IOGH.
DR HAMAP; MF_00146; -; 1.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 156 AA; 17869 MW; 00D57A5A0207B451 CRC64;

Query Match 100.0%; Score 806; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.6e-67;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDVKEGKVVIPPREYALI 60
Db 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDVKEGKVVIPPREYALI 60
QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
Db 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
Db 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
RESULT 2
Q6T180 PRELIMINARY; PRT; 156 AA.
AC Q6T180
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE DUTPase.
OS Pyrococcus woesei.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OC NCBI_TaxID=2262;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=22844507; PubMed=12963343; DOI=10.1016/S1046-5928(03)00108-6;
RA Dabrowski S., Kjaer Ahring B.;
RT "Cloning, expression, and purification of the His6-tagged hyper-
RT thermostable dUTPase from Pyrococcus woesei in Escherichia coli:
RT application in PCR.";
RL Protein Expr. Purif. 31:72-78(2003).
DR EMBL; AY443493; AAR15897.1; -.
DR GO; GO:0046080; P:dutp metabolism; IEA.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
SQ SEQUENCE 156 AA; 17869 MW; 00D57A5A0207B451 CRC64;
Query Match 100.0%; Score 806; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.6e-67;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDVKEGKVVIPPREYALI 60
Db 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDVKEGKVVIPPREYALI 60
QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
Db 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
Db 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
RESULT 3
DCD_PYRHO
ID DCD_PYRHO STANDARD; PRT; 156 AA.
AC O57706;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
DE deaminase).
GN Name=dcdd; OrderedLocusNames=PH1997;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OC NCBI_TaxID=53953;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP000007; BAA31124.1; -.
DR PIR; E71216; E71216.
DR HSSP; Q57972; 10GH.
DR HAMAP; MF_00146; -.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 156 AA; 17871 MW; 96B2C2C50393A985 CRC64;
Query Match 95.3%; Score 768; DB 1; Length 156;
Best Local Similarity 93.6%; Pred. No. 2.6e-63;
Matches 146; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDVKEGKVVIPPREYALI 60
Db 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDVKEGKVVIPPREYALI 60
QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
Db 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
Db 121 RFVQIVFIRLEDPRNPYSGNYQGSTRLVFSKRKKL 156
RESULT 4
DCD_PYRAB
ID DCD_PYRAB STANDARD; PRT; 154 AA.
AC Q9UKS8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
DE deaminase).
GN Name=dcdd; OrderedLocusNames=PYRAB17800; ORFNames=PAB1164;
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OC NCBI_TaxID=29292;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
```



```
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -|- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -|- PATHWAY: De novo synthesis of thymidylate.
CC -|- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ248288; CAB50685.1; -.
DR PIR; G75030; G75030.
DR HSP; Q57872; LOGH.
DR HAMAP; MF_00146; -.
DR InterPro; IPR003232; dCTP deaminase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 154 AA; 17758 MW; C031BEE419094DDB CRC64;

Query Match 87.2%; Score 703; DB 1; Length 154;
Best Local Similarity 85.6%; Pred. No. 2.7e-57;
Matches 131; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVGREAFVKGKLDVEGKVIPPYALI 60
DB 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVGREAFVKGKLDVEGKVIPPYALI 60

QY 61 LTLEIRKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEPVELRYGE 120
DB 61 LTLEIRKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEPVELRYGE 120

QY 121 RFVQIAFLIEGPARNPYRGNGQSTRFLAFSKR 153
DB 121 RFVQIAFLIEGPARNPYRGNGQSTRFLAFSKR 153

RESULT 5
DCD THEAC STANDARD; PRT; 166 AA.
AC Q9HKK0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
DE deaminase).
GN Name=cdcd, orderedLocusNames=Ta0598;
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]_TaxID=2303;
RP SEQUENCE FROM N.A.
RX STRAIN=DSM 1728;
RC MEDLINE=20479972; PubMed=11029001; DOI=10.1038/35035069;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
CC -|- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -|- PATHWAY: De novo synthesis of thymidylate.
CC -|- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL445064; CAC11737.1; ALT_INIT.
DR HSP; Q07199; 1M07.
DR HAMAP; MF_00146; -.
DR InterPro; IPR003232; dCTP deaminase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 166 AA; 18293 MW; 1F4C9467CA3BFBD8 CRC64;

Query Match 31.9%; Score 257.5; DB 1; Length 166;
Best Local Similarity 41.3%; Pred. No. 6.5e-16;
Matches 64; Conservative 22; Mismatches 50; Indels 19; Gaps 5;

QY 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVGREAFVKGKLDVE--KEGKVIP 53
DB 1 MILNDSTIMRMVSDGLLISENFDGCLTPNGYDLRV-----DAIDVEGRQYSFEIG 52

QY 54 PREYALILTLEIRKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEP 113
DB 53 KNVHFLVSTIEILKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEP 113

QY 114 VELRYGERFVQIAFLIEGPARNPYR---GNYQGS 145
DB 112 VNLRRGERIAQIVFVKMIGSAEKPYHIRSGNYQNS 146

RESULT 6
Q684E1 PRELIMINARY; PRT; 172 AA.
ID Q684E1;
AC Q684E1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE dUTPase.
GN Name=dut;
OS Sulfolobus tengchongensis spindle-shaped virus STSV1.
OC Viruses; dsDNA viruses, no RNA stage; Fuselloviridae; Fusellovirus.
OX NCBI_TaxID=285013;
RN [1]
RP SEQUENCE FROM N.A.
RA Xiang X., Chen L., Huang X., She O., Huang L.;
RT "The Sulfolobus tengchongensis spindle-shaped virus STSV1: virus-host
RT interactions and genomic features.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ783769; CAH04216.1; -.
DR InterPro; IPR003232; dCTP deaminase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
SQ SEQUENCE 172 AA; 19974 MW; 0999F5F1488F095 CRC64;

Query Match 30.4%; Score 245; DB 2; Length 172;
Best Local Similarity 37.9%; Pred. No. 9.8e-15;
Matches 55; Conservative 28; Mismatches 48; Indels 14; Gaps 4;

QY 12 ILIEPFSESLQAGYDLRVGRE--AFVKGKLD-----VEKE--GKVVIPREYA 58
DB 17 IKIEPLREDTIRENGVDLRIGNEIVRFKQNRIFDPDKSDIDDFIEKEVGNEFIINPHEHV 76

QY 59 LILTLEIRKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEPVELRY 118
DB 77 LLVTBEYVKLPNDVMAFVNLRSFARLFIPTTVDAGFEQGLTIELV-GSEFPILKLY 135

QY 119 GERFVQIAFLIEGPARNPYRGNYQ 143
DB 136 GMRFIHLIFAKLTLPVEKPYNGYQ 160
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AE000747; AAC07499.1; -;
DR PIR; A70439; A70439.
DR HSSP; Q57872; 10GH.
DR HAWAP; MF_00146; -; 1.
DR InterPro; IPR003232; dCTP_deaminase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
KW COMPLETE proceme; Hydrolyase; Nucleotide metabolism.
SQ SEQUENCE 180 AA; 20544 MW; B2710421A2FA48D6 CRC64;

Query Match 30.1%; Score 243; DB 1; Length 180;
Best Local Similarity 37.7%; Pred. No. 1.6e-14;
Matches 61; Conservative 30; Mismatches 53; Indels 18; Gaps 5;
Qy 1 MLLPDWKIRK-----EILIEPFSESLOPAGYDLRVGRE-AFVKGK-LIDVEKEGKV-- 51
Db 1 MILSDRSIRELIEKGLKVEPEYPSHVQCSLRLGNQIALYEGEVIDVKGTGVRI 60
Qy 52 -----IPPEVALILTLEIRIKLPDDVMGDMKIRSLAREGV-IGSFANVDPGWDGNTL 102
Db 61 LEPEYFDIMPKQLTLATLLEYISLPPYVTAFAVGRSLGLRGLFIENAGWVDAGFEQI 120
Qy 103 TMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQ 144
Db 121 TLELFANDRPIRLYGRNRCQLVFARLDRPPEVYSKYKG 162

RESULT 9
DCD_BACD STANDARD; PRT; 177 AA.
ID DCD BACD
AC Q9AFV3; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase).
GN Name=dcd; OrderedLocusNames=BH0368;
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132; DOI=10.1093/nar/28.21.4317;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -|- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -|- PATHWAY: De novo synthesis of thymidylate.
CC -|- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001508; BAB04087.1; -;
DR PIR; H83695; H83695.
DR HSSP; Q57872; 10GH.

RESULT 7
Q8V9P5 PRELIMINARY; PRT; 158 AA.
ID Q8V9P5
AC Q8V9P5; 2003 (Rel. 40, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Sulfolobus islandicus rod-shaped virus 2.
OC Viruses; dsDNA viruses, no RNA stage; Rudiviridae; Rudivirus.
OX NCBI_TaxID=157899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HVE10/4;
RX MEDLINE=21874820; PubMed=11878892; DOI=10.1006/viro.2001.1190;
RA Peng X., Blum H., She Q., Mallik S., Brugger K., Garrett R.A.,
RA Zillig W., Prangshwilli D.;
RT "Sequences and replication of genomes of the archaeal rudiviruses SIRV1
RT and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some
RT eukaryal viruses.";
RL Virology 291:226-234(2001).
DR EMBL; AJ344259; CAC87298.1; -;
DR HSSP; Q57872; 1PKK.
DR GO; GO:0046080; P:dUTP metabolism; IEA.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
KW Hypothetical protein.
SQ SEQUENCE 158 AA; 18197 MW; 90AEC4BC9D1971B0 CRC64;

Query Match 30.3%; Score 244; DB 2; Length 158;
Best Local Similarity 34.4%; Pred. No. 1.1e-14;
Matches 54; Conservative 37; Mismatches 60; Indels 6; Gaps 2;
Qy 1 MLLPDWKIRK-----EILIEPFSESLOPAGYDLRVGREAFVKGLIDVEKEGKVIPPR 55
Db 1 MILSDRIKRYINLRLVDPLSEDTIRENGVDLKGNIIRIEEMKKEVKDFEIIYPY 60
Qy 56 EYALILTLEIRIKLPDDVMGDMKIRSLAREGVISFANVDPGWDGNTLMLYNASNEPVE 115
Db 61 EHVLTTEYKLPNDIIFAFNLRSTFARKGLLIPTTIVDAGFEGQLTIELV-GSSIPVK 119
Qy 116 LRYGRFVQIAFIRLEGPARNPYRGNYQSTRLAFSK 152
Db 120 LKSGERFHLIFARITVPERPHGKYQKQGVTLAK 156

RESULT 8
DCD_AQUAE STANDARD; PRT; 180 AA.
ID DCD AQUAE
AC Q67539; 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase).
GN Name=dcd; OrderedLocusNames=AQ_1607;
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320; DOI=10.1038/32831;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olsen G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:353-358(1998).
CC -|- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -|- PATHWAY: De novo synthesis of thymidylate.
CC -|- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----

```
DR HAMAP: MF 00146; -: 1.
DR InterPro: IPR003232; dCTP deaminase.
DR InterPro: IPR008180; DeoxyUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP deaminase; 1.
DR Complete proteome; Hydrolase; Nucleotide metabolism.
KW SEQUENCE 177 AA; 19897 MW; D15AE75387847B2B CRC64;

Query Match      30.1%; Score 242.5; DB 1; Length 177;
Best Local Similarity 37.2%; Pred. No. 1.7e-14;
Matches 58; Conservative 30; Mismatches 47; Indels 21; Gaps 6;

QY 10 KEILIEPFSEESLOPAGYDLRVGREAFV-----KGKLIDVEK-----EGKVVIP 54
DB 15 KEILIEPFSEESLOPAGYDLRVGREAFV-----KGKLIDVEK-----EGKVVIP 73
QY 55 REYALITLERIKLPDDVMDMKIRSLAREGV-IGSFAWVDPGWDGNLTMLYNASNEP 113
DB 74 HTPFLATTMETVKLPNHLTAFAVEGRSSVGLGLGFIQAGWVDGFGNGQITLLELFNANRLP 133
QY 114 VELRYGERFVQIAFIRLEGPARNPYRGNY---QGST 146
DB 134 IELPIGRICQLVFAEVTGEVA-PYQGYLFOKGAT 168

RESULT 10
DCD ACIAM STANDARD; PRT; 173 AA.
AC Q02103;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase).
GN Name=dcd;
OS Acidianus ambivalens (Desulfohalobium ambivalens).
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Acidianus.
OX NCBI_TaxID=2283;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lei 10 / DSM 3772;
RX MEDLINE=93065206; PubMed=1437556;
RA Kletzin A.;
RT "Molecular characterisation of a DNA ligase gene of the extremely thermophilic archaeon Desulfohalobium ambivalens shows close phylogenetic relationship to eukaryotic ligases.";
RL Nucleic Acids Res. 20:5389-5396(1992).
RN [2]
RP SIMILARITY.
RX MEDLINE=95206934; PubMed=7899076;
RA Ouzounis C., Kyriades N., Sander C.;
RT "Novel protein families in archaean genomes.";
RL Nucleic Acids Res. 23:565-570(1995).
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X63438; CAA45033.1; -.
DR PIR; S26382; S26382.
DR HSSP; Q57872; IOGH.
DR HAMAP; MF 00146; -: 1.
DR InterPro: IPR003232; dCTP deaminase.
DR InterPro: IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
```

```
DR ProDom: PD004900; dCTP deaminase; 1.
KW SEQUENCE 173 AA; 19858 MW; B4D922503CD4B25A CRC64;

Query Match      30.0%; Score 242; DB 1; Length 173;
Best Local Similarity 35.9%; Pred. No. 1.9e-14;
Matches 56; Conservative 30; Mismatches 56; Indels 14; Gaps 3;

QY 12 ILIEPFSEESLOPAGYDLRVG-----RFAFVKGK-----LIDVEKGVVIPPREYA 58
DB 17 IIVSPLTQDTIRENGVDLRVGGGEIARFKTKTDEIVEDGKDPSPFYETIEKGDEFIYPNEHV 76
QY 59 LILTLERIKLPDDVMDMKIRSLAREGVIGSFAWVDPGWDGNLTMLYNASNEPVELRY 118
DB 77 LIVTEYVKLPDDVMAFVNLRSFARLGLFVPTVIDAGFEGQTLTEVL-GSAPPVKIKR 135
QY 119 GERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRK 154
DB 136 GTFRLHLIFARTLTPVENPYHGKYQGQGVTLPPKPK 171

RESULT 11
O71028 PRELIMINARY; PRT; 158 AA.
AC Q71028;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DUTPase (EC 3.6.1.23).
OS Archaeal virus SIRV.
OC Viruses; unclassified viruses.
OX NCBI_TaxID=66287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98165772; PubMed=9497317; DOI=10.1074/jbc.273.11.6024;
RA Prangishvili D., Klenk H.P., Jakobs G., Schmiechen A., Hanselmann C., Holz I., Zillig W.;
RT "Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV.";
RL J. Biol. Chem. 273:6024-6029(1998).
DR EMBL; AF022221; AAC15873.1; -.
DR HSSP; Q57872; LPKK.
DR GO; GO:0004170; F:dUTP diphosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0046080; P:dUTP metabolism; IEA.
DR InterPro: IPR003232; dCTP deaminase.
DR InterPro: IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP deaminase; 1.
KW Hydrolase.
SQ SEQUENCE 158 AA; 17906 MW; 5BA47765E0889190 CRC64;

Query Match      29.4%; Score 237; DB 2; Length 158;
Best Local Similarity 34.4%; Pred. No. 4.9e-14;
Matches 54; Conservative 36; Mismatches 61; Indels 6; Gaps 2;

QY 1 MLLPDWKIR-----KEILIEPFSEESLOPAGYDLRVGREAFVKGLIDVEKGVVIPPR 55
DB 1 MILSDRDIKTYNSKKLVINPLSEDTIRENGVDLKGNEIVRIKENMEKEVGDEFIYPN 60
QY 56 EVALITLERIKLPDDVMDMKIRSLAREGVIGSFAWVDPGWDGNLTMLYNASNEPVE 115
DB 61 EHVLTITKEYIKLSNDIAFACNLRLSTFARKGLLIPPTVIDAGFEGQTLTELV-GSSIPVK 119
QY 116 LAYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSK 152
DB 120 LKSGERFLHLIFARTLTPVEKPYNGKYQKQKGVTLAK 156

RESULT 12
O777W3 PRELIMINARY; PRT; 158 AA.
ID Q777W3
AC Q777W3;
```


Search completed: July 2, 2005, 12:24:39
Job time : 32.2558 secs

```

RESULT 15
DCD_SULTO      STANDARD;          PRT;    183 AA.
ID -DCD_SULTO
AC Q976G3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
DE deaminase).
DN Name=dcd; OrderedLocusNames=ST0226;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kougai H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
CC -1- CATALYTIC ACTIVITY: dCTP + H2O = dUMP + NH(3).
CC -1- PATHWAY: De novo synthesis of thymidylate.
CC -1- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on use
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; AP000981; BAB65184.1; ALT_INIT.
DR HSSP; Q57872; LOGH.
DR HAMAP; MF 00146; -; 1.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUrase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
DR K0 Complete proteome; Hydrolase; Nucleotide metabolism.
KW SEQUENCE 183 AA; 21055 MW; B3081C863DD71A47 CRC64;
SQ
Query Match 28.9%; Score 233; DB 1; Length 183;
Best Local Similarity 37.7%; Pred. No. 1.4e-13;
Matches 58; Conservative 27; Mismatches 55; Indels 14; Gaps 4;

Qy 12 ILIIPFSESLQAPGYDLRVGRE--AFVK-CKLID-----VEKEGKVIPPREYA 58
Db 17 IKIDPLREDTVRENGVDLVRGELTARFKTDKVPDNDPDAFFKIEEFLIQIEHV 76
Qy 59 LILTLERIKLPDDVGMQDKIRSSLAREGVIGSFAWVDPGDWGNITLMLYNASNEFVELRY 118
Db 77 LLTTEEYIELNDVWAFNWLRSFARLGLFTPTTIVDAGFKGQITIEV-GSSFPVLLR 135
Qy 119 GERFVQIAIRLEGPANPNRYNGYQSGTRLAFSK 152
Db 136 GTRERHLIFARTLSFVEHPHYQGYKGYOGKGVTLPK 169

```

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 2, 2005, 09:02:52 ; Search time 163 Seconds
(without alignments)
1566.007 Million cell updates/sec

Title: US-08-957-709A-71

Perfect score: 806

Sequence: 1 MLLPWKIRKEILBPFSEE.....PYRGNVQSTRLAFSKRKL 156

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US08957709/runat_01072005_154632_26031/app.query.fasta_1.327
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=blosum62 -TRANS=human0.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US08957709 @CNG 1 1 177 @runat 01072005_154632_26031 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	806	100.0	471	4	US-09-399-003-70 Sequence 70, Appl
2	543	67.4	740	3	US-08-822-774-42 Sequence 42, Appl
3	543	67.4	740	3	US-09-632-711-42 Sequence 42, Appl
4	543	67.4	740	3	US-09-632-703B-42 Sequence 42, Appl
5	543	67.4	740	3	US-09-632-702-42 Sequence 42, Appl
6	543	67.4	740	4	US-09-399-003-42 Sequence 42, Appl
7	194	24.1	4403765	3	US-09-103-840A-2 Sequence 2, Appl
8	194	24.1	4411529	3	US-09-103-840A-1 Sequence 1, Appl
9	185.5	23.0	594	4	US-09-328-352-539 Sequence 539, App
10	174	21.6	640681	4	US-09-790-988-1 Sequence 1, Appl
11	172	21.3	609	4	US-09-252-991A-1091 Sequence 1091, Ap
12	168.5	20.9	579	4	US-09-540-236-1761 Sequence 1761, Ap

13	168.5	20.9	615	4	US-09-543-681A-891 Sequence 891, App
14	168.5	20.9	8947	4	US-09-596-002-34 Sequence 34, Appl
15	164	20.3	654	4	US-09-489-039A-1128 Sequence 1128, Ap
16	160	19.9	129	3	US-08-822-774-40 Sequence 40, Appl
17	160	19.9	129	3	US-09-632-711-40 Sequence 40, Appl
18	160	19.9	129	3	US-09-632-703B-40 Sequence 40, Appl
19	160	19.9	129	3	US-09-632-702-40 Sequence 40, Appl
20	160	19.9	129	4	US-09-399-003-40 Sequence 40, Appl
21	154	19.1	1230025	4	US-09-198-452A-1 Sequence 1, Appl
22	154	19.1	1230230	4	US-09-438-185A-1 Sequence 1, Appl
23	154	19.1	1664976	4	US-08-916-421B-1 Sequence 1, Appl
24	154	19.1	1664976	4	US-09-692-570-1 Sequence 1, Appl
25	149.5	18.5	1830121	4	US-09-557-884-1 Sequence 1, Appl
26	149.5	18.5	1830121	4	US-09-643-990A-1 Sequence 1, Appl
27	134.5	16.7	534	3	US-09-199-637A-246 Sequence 246, App
28	134.5	16.7	42235	3	US-09-199-637A-1 Sequence 1, Appl
29	132.5	16.4	501	3	US-09-199-637A-244 Sequence 244, App
30	123	15.3	84	3	US-08-822-774-36 Sequence 36, Appl
31	123	15.3	84	3	US-09-632-711-36 Sequence 36, Appl
32	123	15.3	84	3	US-09-632-703B-36 Sequence 36, Appl
33	123	15.3	84	3	US-09-632-702-36 Sequence 36, Appl
34	123	15.3	84	4	US-09-399-003-36 Sequence 36, Appl
35	121	15.0	444	4	US-09-583-110-1776 Sequence 1776, Ap
36	121	15.0	477	4	US-09-107-433-924 Sequence 924, App
37	118	14.6	456	4	US-09-902-540-3295 Sequence 3295, Ap
38	118	14.6	11199	4	US-09-902-540-1017 Sequence 1017, Ap
39	117.5	14.6	1664976	4	US-08-916-421B-1 Sequence 1, Appl
40	117.5	14.6	1664976	4	US-09-692-570-1 Sequence 1, Appl
41	117	14.5	10223	3	US-08-961-527-73 Sequence 73, Appl
42	112	13.9	546	3	US-09-199-637A-238 Sequence 238, App
43	111	13.8	960	2	US-08-824-405-3 Sequence 3, Appl
44	111	13.8	1037	2	US-08-824-405-1 Sequence 1, Appl
45	111	13.8	1326	4	US-09-949-016-5063 Sequence 5063, Ap

ALIGNMENTS

RESULT 1

US-09-399-003-70
; Sequence 70, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Högrefe, Holly
; APPLICANT: Hansen, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complex
; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
US-09-399-003-70

Alignment Scores:
Pred. No.: 1 66e-108 Length: 471
Score: 806.00 Matches: 156
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-08-957-709A-71 (1-156) x US-09-399-003-70 (1-471)

QY 1 MetLeuLeuProAspTyrLysIleArgLysGluIleLeuLeuGluProPheSerGluGlu 20
DB 1 ATGCTACTTCAGACTGGAAATCAGAAAGAAATACCTTATAGAGCCATTTCTGAAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAACACGACGAGGTATGACCTCAGAGTGGCGAGAGAGGCTTTTGTAAAGGGGAAA 120
QY 41 LeuIleAspValGluLysGluLysValValIleProProArgGluTyrAlaLeuIle 60
DB 121 TTAATCGACGTGGAAGAGGAAGAAAGTCGTATTCTCCAAAGGAATACGCCCTTAATC 180
QY 61 LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleArgSer 80
DB 181 CTAACCTTCGAGAGGATAAGTTGCCGACGATGTTATGGGGATATGAAGATAAGGAGC 240
QY 91 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTATTGGTTCTTTTGGTTCGCCAGGATGGATGGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACACTAATGCTCTACAAATGCCCTCAATGAACCTGTGCAATTAAGATATGGAGAG 360
QY 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
DB 361 AGATTGTGCAGATCGCATTTATAGGCTAGAGGGTCCGGCAGAAACCTTACAGAGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLeu 156
DB 421 AACTATCAGGCGAGCACAGGTTAGCGTTTCCAAAGAGAAAGAACTC 468

RESULT 2

US-08-822-774-42
; Sequence 42, Application US/08822774
; Patent No. 6183997

GENERAL INFORMATION:

; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSES: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997

CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO

; ANTI-SENSE: NO
US-08-822-774-42

Alignment Scores:

Score: 1.23e-69 Length: 740
Pred. No.: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 3 Gaps: 0

US-08-957-709A-71 (1-156) x US-08-822-774-42 (1-740)

QY 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
DB 2 CCTCAAGGGAATACGCCCTTAATCTTAACCTCGAGAGGATAAAAGTTGCCCGCAGATGTT 61
QY 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGGGGTATTGGTTCTTTTGTCT 121
QY 93 TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTGTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTCAATGCCCTCAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
DB 182 CTGTCTGAAATTAAGATATGGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAGGTTAGCGTTTTCAAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAGAAACTC 313

RESULT 3

US-09-632-711-42
; Sequence 42, Application US/09632711
; Patent No. 6333165

GENERAL INFORMATION:

; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSES: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.

REGISTRATION NUMBER: 36,576

REFERENCE/DOCKET NUMBER: 1486/43163

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-8800

TELEFAX: (202) 628-8844

INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-711-42

Alignment Scores:
Pred. No.: 1,23e-69 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 3 Gaps: 0

US-08-957-709A-71 (1-156) x US-09-632-711-42 (1-740)

QY 53 ProProArgGluTyrAlaLeuLeuThrLeuGluArgLysLeuProAspVal 72
DB 2 CTCCTCAAGGATACGCTTATCTTACCTCGAGAGGATAAAGTTGCCGACGATGT 61
QY 73 MetGlyAspMetLysIleArgSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATGAAGATAAGGACGATTTAGCAAGAGAGGGGTATTGCTTTTGTCT 121
QY 93 TrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTTGACCCAGGATGGATGGAATCTTAACACTAATGCTCTACAAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
DB 182 CTGTGCAATTAGATATGAGAGAGATTTGTCAGATCGCATTTATAGGCTTAGAGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCTTACAGAGGAACTATCAGGGGAGCACAAAGTTAGCGTTTCAAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAGAAACTC 313

RESULT 4

US-09-632-703B-42
Sequence 42, Application US/09632703B
Patent No. 6379553
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,703B
FILING DATE: 24-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/822,774

FILING DATE: 21-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: BARKER, M. Paul
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 4121.0116-04
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 849-6613
TELEFAX: (650) 849-6686
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-703B-42

Alignment Scores:
Pred. No.: 1,23e-69 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 3 Gaps: 0

US-08-957-709A-71 (1-156) x US-09-632-703B-42 (1-740)

QY 53 ProProArgGluTyrAlaLeuLeuThrLeuGluArgLysLeuProAspVal 72
DB 2 CTCCTCAAGGATACGCTTATCTTACCTCGAGAGGATAAAGTTGCCGACGATGT 61
QY 73 MetGlyAspMetLysIleArgSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATGAAGATAAGGACGATTTAGCAAGAGAGGGGTATTGCTTTTGTCT 121
QY 93 TrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTTGACCCAGGATGGATGGAATCTTAACACTAATGCTCTACAAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
DB 182 CTGTGCAATTAGATATGAGAGAGATTTGTCAGATCGCATTTATAGGCTTAGAGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCTTACAGAGGAACTATCAGGGGAGCACAAAGTTAGCGTTTCAAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAGAAACTC 313

RESULT 5

US-09-632-702-42
Sequence 42, Application US/09632702
Patent No. 6444428
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,702
FILING DATE: 04-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-Mar-1997
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
TELEPHONE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-702-42

Alignment Scores:
Pred. No.: 1,236-69 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 3 Gaps: 0
US-08-957-709A-71 (1-156) x US-09-632-702-42 (1-740)
QY 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLeuLeuProAspVal 72
DB 2 CCTCAAGGGAATAGCCCTTAATCCTAACCTCGAGAGGATAAAGTTGCCCGACGATGT 61
QY 73 MetGlyAspMetLysIleArgSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATGAAGATAAGGAGCAGTTTAGCAGAGAGGGGTTATTGGTTCTTTTGGCT 121
QY 93 TrpValaspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACAATGCCCTCAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGly 132
DB 182 CCTGTCGAATTAAAGATATGGAGAGAGATTGTGCAGATCGCATTTATTAAGGCTAGAGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAGGTTAGCGTTTCAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAGAAACTC 313

RESULT 6

US-09-399-003-42
Sequence 42, Application US/09399003
Patent No. 6734293
GENERAL INFORMATION:
APPLICANT: Hogeefe, Holly
APPLICANT: Hansen, Connie J
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
FILE REFERENCE: 4121.0116-02

CURRENT APPLICATION NUMBER: US/09/399,003
CURRENT FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: PCT/ US98/05497
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: US 08/957,709
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: US 08/822,774
PRIOR FILING DATE: 1997-03-21
NUMBER OF SEQ ID NOS: 89
SOFTWARE: PatentIn version 3.1
SEQ ID NO 42
LENGTH: 740
TYPE: DNA
ORGANISM: Pyrococcus furiosus
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(740)
OTHER INFORMATION: "n" is a, t, g, or c
US-09-399-003-42

Alignment Scores:
Pred. No.: 1,236-69 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 4 Gaps: 0

US-08-957-709A-71 (1-156) x US-09-399-003-42 (1-740)

QY 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLeuLeuProAspVal 72
DB 2 CCTCAAGGGAATAGCCCTTAATCCTAACCTCGAGAGGATAAAGTTGCCCGACGATGT 61
QY 73 MetGlyAspMetLysIleArgSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATGAAGATAAGGAGCAGTTTAGCAGAGAGGGGTTATTGGTTCTTTTGGCT 121
QY 93 TrpValaspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACAATGCCCTCAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGly 132
DB 182 CCTGTCGAATTAAAGATATGGAGAGAGATTGTGCAGATCGCATTTATTAAGGCTAGAGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAGGTTAGCGTTTCAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAGAAACTC 313

RESULT 7

US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Alignment Scores:
Pred. No.: 1.6e-12 Length: 4403765
Score: 194.00 Matches: 50
Percent Similarity: 49.70% Conservative: 34
Best Local Similarity: 29.59% Mismatches: 61
Query Match: 24.07% Indels: 24
DB: 3 Gaps: 3

US-08-957-709A-71 (1-156) x US-09-103-840A-2 (1-4403765)

QY 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeu-----lleGlu 15
DB 388635 GTGCTGCTCTCCGATCGTATCTTCGGCCGAGATCTCTCCGGCGGTTGGGGATCGAC 388694
QY 16 PropheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgVal-----31
DB 388695 CGGTTCCGACGACACCTGCTCCAGCGCTCCAGCATCGAGTCCGGCTCGATTGCTTTGTTT 388754
QY 32 -----GlyArgGluAlaPheValLys 38
DB 388755 CGGGTGTTCACACACTCGCTACACCCATCGACCCCGCCAGCAGCAGCAGCTG 388814
QY 39 GlyLysLeuIleAspValGluLysGluGlyLysValIleProProArgGluTyrAla 58
DB 388815 ACCAGCTTGGTCAACCGCTGACCGGGGAACCTTCGTGTTGCCCGCGGCGAATTCGTG 388874
QY 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIle 78
DB 388875 CTCGGCTCGACGCTGGAGCTTTTCACTTCGCCGACCACTCCCGCGGCTGAGAGC 388934
QY 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
DB 388935 AAGTCTTCGTTGGCGCGCTGGCGCTGCTGACGCATTCACCGCGGCTTCATCGATCCT 388994
QY 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
DB 388995 GCGTTCCAGCGGTTCACATCACCCTGGAGCTATCCAACTGTCGATGCTGCGCATCTTTG 389054
QY 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
DB 389055 TGCCCGCGCATGAATATCGTTCAGCTGCTGATGTTGGCTGACCAAGCCGTCGAGCAT 389114
QY 137 ProTyrArgGlyAsnTyrGlnGlySer 145
DB 389115 CCTACGGCAGTTCCTCCGGCGGGGTGCG 389141

RESULT 8

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; FILE REFERENCE: TUBERCULOSIS

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 4411529

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; OTHER INFORMATION: H37RV

US-09-103-840A-1

Alignment Scores:

Pred. No.: 1.6e-12 Length: 4411529
Score: 194.00 Matches: 50
Percent Similarity: 49.70% Conservative: 34
Best Local Similarity: 29.59% Mismatches: 61
Query Match: 24.07% Indels: 24
DB: 3 Gaps: 3

US-08-957-709A-71 (1-156) x US-09-103-840A-1 (1-4411529)

QY 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeu-----lleGlu 15
DB 388580 GTGCTGCTCTCCGATCGTATCTTCGGCCGAGATCTCTCCGGCGGTTGGGGATCGAC 388639
QY 16 PropheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgVal-----31
DB 388640 CGGTTCCGACGACACCTGCTCCAGCGCTCCAGCATCGAGTCCGGCTCGATTGCTTTGTTT 388699
QY 32 -----GlyArgGluAlaPheValLys 38
DB 388700 CGGGTGTTCACACACTCGCTACACCCATCGACCCCGCCAGCAGCAGCTG 388759
QY 39 GlyLysLeuIleAspValGluLysGluGlyLysValIleProProArgGluTyrAla 58
DB 388760 ACCAGCTTGGTCAACCGCTGACCGGGGAACCTTCGTGTTGCCCGCGGCGAATTCGTG 388819
QY 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIle 78
DB 388820 CTCGGCTCGACGCTGGAGCTTTTCACTTCGCCGACCACTCCCGCGGCGGCTGAGAGC 388879
QY 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
DB 388880 AAGTCTTCGTTGGCGCGCTGGCGCTGCTGACGCATTCACCGCGGCTTCATCGATCCT 388939
QY 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
DB 388940 GCGTTCCAGCGGTTCACATCACCCTGGAGCTATCCAACTGTCGATGCTGCGCATCTTTG 388999
QY 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
DB 389000 TGCCCGCGCATGAATATCGTTCAGCTGCTGATGTTGGCTGACCAAGCCGTCGAGCAT 389059
QY 137 ProTyrArgGlyAsnTyrGlnGlySer 145
DB 389060 CCTACGGCAGTTCCTCCGGCGGGGTGCG 389086

RESULT 9

US-09-328-352-539

; Sequence 539, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 539

; LENGTH: 594

; TYPE: DNA

; ORGANISM: Acinetobacter baumannii

US-09-328-352-539

Alignment Scores:

Pred. No.: 2.26e-17 Length: 594
Score: 185.50 Matches: 48
Percent Similarity: 45.14% Conservative: 31
Best Local Similarity: 27.43% Mismatches: 63
Query Match: 23.01% Indels: 33

DB: 4 Gaps: 4

US-08-957-709A-71 (1-156) x US-09-328-352-539 (1-594)

Qy 10 LysGluIleLeuIleGluProPheSerGluGluSerLeuGln----- 23
|||
Db 67 AAACACGGCATGATTGAACCTTATGCAGAGATCAAGTCGCTTTGTATAAGAAATGGCGAA 126
-----ProAlaGlyTyrAspLeuArgValGlyArgGlu--- 34
Qy 24 ----- 186
Db 127 AAATTGATTTCACGGGGTCTCTAGCTATGTTATGACGTACGCTGCGCGCGGAATTT 186
Qy 35 -----AlaPheValIysGlyLys 40
Db 187 AAAGTTTTCACTAACGTACATCTGCAATTGTTGATCCAAAAATTTTCGATGAAAAAGT 246
Qy 41 LeuIleAspValGluGlyGluLysValValIleProProArgGluTyrAlaLeuIle 60
|||
Db 247 TTTATCGATATTGAGTCTGAC---GTTTGTATTATTCGCGCTAACTCATTTGCTCTAGCG 303
Qy 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
|||
Db 304 CGTAAATGTAATTTCCGTATTCACGTAATGTTTGGACTGTGCTGCTGTAATAATCA 363
Qy 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100
|||
Db 364 ACTTATGCACGTTCGGGTATTATTGTAATGTCACCTCTCTAGAGCCAGAAATGGGAAGG 423
Qy 101 AnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
|||
Db 424 CACGTTACTTTAGAAATTTCTAATAACCAAGAACTTGCCTGCGCGTATTATTATGCGGGTGA 483
Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArg--- 139
Db 484 GGTGTAGCACAAATGCTATTTTTCGAAAGTGATGAAGTATGCGAAACCTCTTATAAAGAT 543
Qy 140 -----GlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 544 CGTGTGTGTAATACCAAGGCCAAACTGGGGTTACACTTCCTAAG 588

RESULT 10

US-09-790-988-1

Sequence 1, Application US/09790988

Patent No. 6632935

GENERAL INFORMATION:

APPLICANT: SHIGENOBU, SHUII

APPLICANT: WATANABE, HIDEMI

APPLICANT: HATTORI, NASHIRA

APPLICANT: SAKAKI, YOSHIYUKI

TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS

FILE REFERENCE: 081356/0159

CURRENT APPLICATION NUMBER: US/09/790,988

CURRENT FILING DATE: 2001-02-23

PRIOR APPLICATION NUMBER: JP2000-107160

PRIOR FILING DATE: 2000-04-07

NUMBER OF SEQ ID NOS: 7

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 640681

TYPE: DNA

ORGANISM: Buchnera sp.

US-09-790-988-1

Alignment Scores:

Pred. No.: 6,568-11 Length: 640681

Score: 174.00 Matches: 48

Percent Similarity: 47.56% Conservative: 30

Best Local Similarity: 29.27% Mismatches: 58

Query Match: 21.55% Indels: 28

DB: 4 Gaps: 5

US-08-957-709A-71 (1-156) x US-09-790-988-1 (1-640681)

Qy 3 LeuProAspTrpLysIleArgLysGluIleLeuLeuGluProPheSerGluGluSerLeu 22
|||
Db 113257 ATTGAAGAGTGGTTGGAGAGAAAAGAAATTAATTATAGAACCTTATCTTAATAAACATTA 113316
Qy 23 ---GlnProAlaGlyTyrAspLeuArgValGlyArgGlu-----AlaPheValIys--- 38
|||
Db 113317 ATTAATGCAATTACTGTGATATACACCTTGGTAATAAAATTCGTTTTTTTATGAACAT 113376
Qy 39 ---GlyLysLeuIleAspVal----- 44
Db 113377 ACTGATCTTCATTTGATTTGAGTAATTTCTAAAAATAATCGGTGGATTATCATTAAACGNA 113436
Qy 45 -----GlyLysGluGlyLysValValIleProProArgGlu 56
Db 113437 ATTATGAGTAATGAATAATAATTTTCTAAAGACAGCCATGTTTTTACAACCGAGATCT 113496
Qy 57 TyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMet 76
|||
Db 113497 TTAGTATTATTGTCGACTTTTGAAGATATTAAAAATGCCAAATAATTTAGTGGTTGGTTA 113556
Qy 77 LysIleArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpVal 94
|||
Db 113557 GATCGCGCTTCTTTTAGCTCGTTAGGATTAATGATTCAGGTACCGCTCATCGTATT 113616
Qy 95 AspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProVal 114
|||
Db 113617 GATCAGGTTGGAATGGTAAATTATTTTGTAGAAATGTTTAAATGCAGGAAAAATAACTTTA 113676
Qy 115 GluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAla 134
|||
Db 113677 GTATTACGACCTAAATGAGGATTCGACGACTTAGCTTTGAAGTTCTATCTCAACCCAGTT 113736
Qy 135 ArgAsnProTyr 138
Db 113737 TTACGTCCTCTAT 113748

RESULT 11

US-09-252-991A-1091

Sequence 1091, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 1091

LENGTH: 609

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-1091

Alignment Scores:

Pred. No.: 2,24e-15 Length: 609

Score: 172.00 Matches: 45

Percent Similarity: 45.03% Conservative: 32

Best Local Similarity: 26.32% Mismatches: 62

Query Match: 21.34% Indels: 32

DB: 4 Gaps: 4

US-08-957-709A-71 (1-156) x US-09-252-991A-1091 (1-609)

Qy 13 LeuIleGluProPheSerGluGluSerLeuGlnProAla----- 25
|||
Db 94 ATGATCGAGCCGTTGCTGACGCGCGGAGGAGCGGCGGCGGAGCGCGGCGGATTTCC 153
Qy 26 -----GlyTyrAspLeuArgValGlyArgGlu----- 34

Db 460 GGCTGGTCCGGCTGCATCGTGTGCTTGTAGTTCTACAACTCCGGCAAGTTGCCGCTGGCGCTG 519
 Qy 117 ArgTyr-GlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
 Db 520 CGCCCGGGGATGCCCATCGGCGCGCTGAGCTTTGAACCGTTATCGGGCCCGGCGGCCCGG 579
 Qy 137 ProTyr 138
 Db 580 CCGTAT 585

Search completed: July 2, 2005, 11:27:04
 Job time : 1235 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:34:09 ; Search time 8.65116 Seconds
(without alignments)
1735.006 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 156
Sequence: 1 MLLPDWKIRKEILIEPFSE.....PYRGNVQGSTRLAFSKRKKL 156

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 79.*

- 1: Pirl.*
- 2: Pirl2.*
- 3: Pirl3.*
- 4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	23.7	156	2 E71216	dCTP deaminase (EC
2	22	14.1	154	2 G75030	dCTP deaminase (EC
3	8	5.1	420	2 D84965	dihydrolipoamide S
4	8	5.1	503	2 AB2938	alpha-L-arabinofur
5	8	5.1	503	2 E98344	hypothetical prote
6	7	4.5	133	1 I40398	flagellar protein
7	7	4.5	133	2 E84101	ail protein precurs
8	7	4.5	178	2 A35123	attachment invasio
9	7	4.5	182	2 AI0353	transporter, lyse
10	7	4.5	204	2 AD3411	alkyl hydroperoxid
11	7	4.5	215	2 S74033	phosphoribosylform
12	7	4.5	237	2 C64491	conserved hypothet
13	7	4.5	268	1 A69000	hypothetical prote
14	7	4.5	279	2 A64149	cysteine synthase
15	7	4.5	327	2 A70435	probable transmemb
16	7	4.5	362	2 C81445	homoserine dehydro
17	7	4.5	376	2 T40673	probable porin - A
18	7	4.5	439	2 I39524	allantoinase (EC 3
19	7	4.5	460	1 S48489	two-component sens
20	7	4.5	587	2 H83748	3-hydroxyacyl-CoA
21	7	4.5	661	2 A69252	methyl-accepting c
22	7	4.5	666	2 D82386	18 protein - fowlp
23	7	4.5	682	1 G48563	probable RNA helic
24	7	4.5	684	2 T30652	probable nucleotid
25	7	4.5	782	2 D81281	hypothetical prote
26	7	4.5	874	2 H86167	cell-cycle-depende
27	7	4.5	1017	2 PC4035	type II restrictio
28	7	4.5	1084	2 F95141	type II site-speci
29	7	4.5	1084	2 D98009	

30	7	4.5	1113	2 E64215	hypothetical prote
31	7	4.5	1921	2 T13827	kinesin-73 - fruit
32	6	3.8	34	2 S77646	hypothetical prote
33	6	3.8	35	2 C70256	hypothetical prote
34	6	3.8	37	1 R5EG36	ribosomal protein
35	6	3.8	55	2 H70228	hypothetical prote
36	6	3.8	62	2 I39622	ribosomal protein
37	6	3.8	62	2 AD1899	30S ribosomal prot
38	6	3.8	62	2 H84083	hypothetical prote
39	6	3.8	64	2 T29319	hypothetical prote
40	6	3.8	68	2 G82600	hypothetical prote
41	6	3.8	80	2 A90767	probable terminase
42	6	3.8	89	2 AI1033	hypothetical prote
43	6	3.8	94	1 VUWTEM	embryonic abundant
44	6	3.8	94	2 C97048	hypothetical prote
45	6	3.8	96	2 B82436	conserved hypothet

ALIGNMENTS

RESULT 1

E71216
dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: E71216
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: E71216
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-156 <RAW>
A;Cross-references: UNIPROT:O57706; GB:AP000007; NID:G3236134; PIDN:BAA31124.1; PID:G32
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1997
C;Keywords: hydrolase

Query Match 23.7%; Score 37; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLPDWKIRKEILIEPFSESLQPGYDLRVGREAFTV 37
|||||
Db 1 MLLPDWKIRKEILIEPFSESLQPGYDLRVGREAFTV 37

RESULT 2

G75030
dCTP deaminase (EC 3.5.4.13) PAB1164 [similarity] - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: G75030
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A;Reference number: A75001
A;Accession: G75030
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-154 <RAW>
A;Cross-references: UNIPROT:Q9UXS8; GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CAB506
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: dcd; PAB1164
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

```

Query Match          14.1%; Score 22; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 4.9e-15; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;

QY 89 GSFANVDPGWDGNLTMLYNAS 110
|||||
DB 89 GSFANVDPGWDGNLTMLYNAS 110

RESULT 3
DB4965
dihydrolipoamide S-succinyltransferase (EC 2.3.1.61) [imported] - Buchnera sp. (strain A
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: DB4965
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A:Reference number: AB4930; MUID:20445173; PMID:10993077
A:Accession: DB4965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: sucB; BU303
C:Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology
C:Keywords: acyltransferase; coenzyme A

Query Match          5.1%; Score 8; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 3.9; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 45 EKEGKVI 52
|||||
DB 62 EKEGKVI 69

RESULT 4
AB2938
alpha-L-arabinofuranosidase Atu3104 [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AB2938
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2938
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-503 <KUR>
A:Cross-references: UNIPROT:Q8UBB2; GB:AE008689; PIDN:AAL43920.1; PID:gl7741471; GSPDB:C
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3104
A:Map position: linear chromosome

Query Match          5.1%; Score 8; DB 2; Length 503;
Best Local Similarity 100.0%; Pred. No. 4.6; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 25 AGYDLRVG 32
|||||
DB 455 AGYDLRVG 462

RESULT 5

```

```

E98344
hypothetical protein AGR_L3408 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: E98344
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E98344
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-503 <KUR>
A:Cross-references: UNIPROT:Q8UBB2; GB:AE007870; PIDN:AAK90279.1; PID:gl5160302; GSPDB:C
C:Genetics:
A:Gene: AGR_L3408
A:Map position: linear chromosome

Query Match          5.1%; Score 8; DB 2; Length 503;
Best Local Similarity 100.0%; Pred. No. 4.6; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 25 AGYDLRVG 32
|||||
DB 455 AGYDLRVG 462

RESULT 6
I40398
flagellar protein flis - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: I40398; E69625
R:Chen, L.; Helmann, J.D.
J. Bacteriol. 176, 3093-3101, 1994
A:Title: The Bacillus subtilis sigma D-dependent operon encoding the flagellar proteins
A:Reference number: I40396; MUID:94252974; PMID:8195064
A:Accession: I40398
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-133 <RES>
A:Cross-references: UNIPROT:P39739; EMBL:Z31376; NID:g499379; PIDN:CAA83249.1; PID:g4993
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berlet
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Frits, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapido, A.; Lardinola,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maguda, S.; Maue
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaite, Y.; Sato, T.; Scanlon,
A:Authors: Schleith, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serio
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: AG9580; MUID:98044033; PMID:9384377
A:Accession: E69625
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-133 <KUN>
A:Cross-references: GB:Z99122; GB:AL009126; NID:g2636029; PIDN:CAB15550.1; PID:g2636059
A:Experimental source: strain 168
C:Genetics:
A:Gene: flis
C:Superfamily: flagellar protein flis

Query Match          4.5%; Score 7; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 16; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

```

QY 102 LTLMLYN 108
Db 23 LTLMLYN 29

RESULT 7

E84101
flagellar protein flis [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: E84101
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: E84101
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-133 <STO>
A:Cross-references: UNIPROT:Q9K6W1; GB:AP001519; GB:BA000004; NID:gl0176109; PIDN:BA073
A:Experimental source: strain C-125
C:Genetics:
A:Gene: flis
C:Superfamily: flagellar protein flis

Query Match 4.5%; Score 7; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LTLMLYN 108
Db 23 LTLMLYN 29

RESULT 8 *

A35123
all protein precursor - Yersinia enterocolitica
C:Species: Yersinia enterocolitica
C:Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 09-Jul-2004
C:Accession: A35123
R:Miller, V.L.; Bliska, J.B.; Falkow, S.
J. Bacteriol. 172, 1062-1069, 1990
A:Title: Nucleotide sequence of the Yersinia enterocolitica Ail gene and characterization
A:Reference number: A35123; MUID:90130261; PMID:1688838
A:Accession: A35123
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-178 <MIL>
A:Cross-references: UNIPROT:P16454; GB:M29945; NID:gl55437; PIDN:AAA88694.1; PID:gl55438
C:Superfamily: phage lambda membrane protein lom

Query Match 4.5%; Score 7; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 GVIGSFA 92
Db 65 GVIGSFA 71

RESULT 9

A10353
attachment invasion locus protein [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: A10353
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
Geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: A10353

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <KUR>
A:Cross-references: UNIPROT:Q8ZCR3; GB:AL590842; PIDN:CAC92156.1; PID:gl5980871; GSPDB:
C:Genetics:
A:Gene: ail
C:Superfamily: phage lambda membrane protein lom

Query Match 4.5%; Score 7; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 GVIGSFA 92
Db 68 GVIGSFA 74

RESULT 10

AD3411
transporter, lyse family BME11274 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AD3411
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella meliten
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3411
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <KUR>
A:Cross-references: UNIPROT:Q8YG87; GB:AE008917; PIDN:AAL52455.1; PID:gl7983261; GSPDB:
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11274
A:Map position: 1
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 4.5%; Score 7; DB 2; Length 204;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RLEGPAR 135
Db 138 RLEGPAR 144

RESULT 11 *

S74033
alkyl hydroperoxide reductase homolog - Sulfolobus solfataricus
N:Alternate names: protein c0215
C:Species: Sulfolobus solfataricus
C:Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 09-Jul-2004
C:Accession: S74033
R:Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.
Mol. Microbiol. 22, 175-191, 1996
A:Title: Organizational characteristics and information content of an archaeal genome:
A:Reference number: S73076; MUID:97055432; PMID:8899719
A:Accession: S74033
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-215 <SEN>
A:Cross-references: UNIPROT:P95895; EMBL:Y08256; NID:gl707679; PIDN:CAA69447.1; PID:gl7
A:Experimental source: strain P2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996
C:Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase C22 pro

Query Match 4.5%; Score 7; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 RIKLPDD 71

```

Db      24 RIKLPDD 30
|||||
RESULT 12
C64491
phosphoribosylformimino-5-aminomidazole carboxamide ribotide isomerase - Methanococcus
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: C64491
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
raon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: C64491
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-237 <BUL>
A:Cross-references: UNIPROT:Q58927; GB:U67594; GB:L77117; NID:gl592160; PIDN:AAB99553.1;
C:Genetics:
C:Superfamily: N-(5'-phospho-D-ribose)lformimino)-5-amino-1-(5''-phosphoribosyl)-4-imidaz
C:Map position: FOR1508884-1509597
Query Match      4.5%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      46 KEGKVI 52
|||||
Db      131 KEGKVI 137
|||||
RESULT 13
A69000
conserved hypothetical protein MTH1 - Methanobacterium thermoautotrophicum (strain Delta
C:Species: Methanobacterium thermoautotrophicum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A69000
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: A69000
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-268 <MTH>
A:Cross-references: UNIPROT:Q26109; GB:AE000795; GB:AE000666; NID:g2621036; PIDN:AAB8451
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1
C:Superfamily: conserved hypothetical protein MTH1
Query Match      4.5%; Score 7; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      10 KEILIEP 16
|||||
Db      149 KEILIEP 155
|||||
RESULT 14
A64149
hypothetical protein HI0345 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C:Accession: A64149
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A

```

```

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: A64149
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-279 <TIGR>
A:Cross-references: UNIPROT:P44652; GB:U32719; GB:L42023; NID:gl573310; PIDN:AAC22006.1;
A:Note: best homolog was a hypothetical protein from Escherichia coli
C:Superfamily: conserved hypothetical protein HI0345; ferredoxin 2[4Fe-4S] homology
F:65-130/Domain: ferredoxin 2[4Fe-4S] homology <FER>

```

```

Query Match      4.5%; Score 7; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      81 SLAREGV 87
|||||
Db      48 SLAREGV 54
|||||

```

```

RESULT 15
A70435
cysteine synthase (EC 4.2.99.8) cysM [similarity] - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: A70435
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.B.; Ov
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: A70435
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-327 <AQF>
A:Cross-references: UNIPROT:O67507; GB:AE000745; NID:g2983907; PIDN:AAC07459.1; PID:g298
A:Experimental source: strain VF5
C:Genetics:
A:Gene: cysM
C:Superfamily: threonine dehydratase
C:Keywords: carbon-oxygen lyase; phosphoprotein; pyridoxal phosphate
F:65/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

```

```

Query Match      4.5%; Score 7; DB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      46 KEGKVI 52
|||||
Db      83 KEGKVI 89
|||||

```

```

Search completed: July 2, 2005, 12:38:51
Job time : 11.6512 secs

```

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	102	25.3	402	2	Q8X250	pyrococcus
2	72	17.9	401	2	O59114	pyrococcus
3	60	14.9	401	2	Q9V0S3	pyrococcus
4	12	3.0	394	2	Q9X213	thermotoga
5	12	3.0	410	2	Q83NK7	tropheryma
6	12	3.0	417	2	Q83MW7	tropheryma
7	11	2.7	403	1	DFP METJTA	m coenzyme
8	10	2.5	86	2	Q971S3	pyrolobus
9	10	2.5	394	2	Q9UT98	neisseria m
10	10	2.5	394	2	Q9UTB7	neisseria m
11	10	2.5	399	2	Q9RQH7	listeria mo
12	10	2.5	399	2	Q8Y674	listeria mo
13	10	2.5	399	2	Q92AI3	listeria in
14	10	2.5	399	2	Q71VJ1	listeria mo
15	10	2.5	412	2	Q976C4	pyrolobus
16	9	2.2	382	2	Q9RH70	methanobact
17	9	2.2	404	2	Q28628	archaeoglob
18	9	2.2	409	2	Q82UM0	nitrosomona
19	9	2.2	414	2	Q6NH19	corynebacte
20	9	2.2	431	2	Q8TRA0	methanosarc
21	9	2.2	1185	2	Q94H40	oryza sativ
22	8	2.0	118	2	Q720W6	listeria mo
23	8	2.0	127	2	Q9XUT2	caenorhabdi
24	8	2.0	142	2	Q9RI41	mus musculu
25	8	2.0	169	2	Q9ZSW7	hamamelis v
26	8	2.0	178	2	Q72JB0	thermus the
27	8	2.0	206	2	Q70629	anopheles g
28	8	2.0	210	2	Q91596	xenopus lae
29	8	2.0	210	2	Q6GNQ5	xenopus lae
30	8	2.0	220	2	Q6L0U4	photobacter
31	8	2.0	227	1	GPNA RAEDU	haemophilus

DE Hypothetical protein PH1444.
GN OrderedLocusNames=PH1444;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
Maeuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000006; BAA30551.1; -;
DR PIR; G71018; G71018.
DR HSP; F30197; IG63.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DRP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
SQ Complete proteome; Hypothetical protein.
KW SEQUENCE 401 AA; 44624 MW; 26CBC5F523AE02C8 CRC64;

Query Match 17.9%; Score 72; DB 2; Length 401;
Best Local Similarity 100.0%; Pred.No. 3.4e-61; Mismatches 0; Indels 0; Gaps 0;
Matches 72; Conservative 0;

QY 68 TGNPVTITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTTAPP 127
DB |||||||
69 TGNPVTITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTTAPP 128

QY 128 HIPIMIAPAMHE 139
DB |||||||
129 HIPIMIAPAMHE 140

RESULT 3
ID Q9V0S3 PRELIMINARY; PRT; 401 AA.
AC Q9V0S3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Dfp DNA/pantothenate metabolism flavoprotein.
GN ORFNames=PAB1897;
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
Van der Oost J., Weisenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
DR EMBL; AJ248285; CAB49630.1; -;
DR PIR; E75114; E75114.
DR HSP; P30197; IG63.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.

DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 401 AA; 44290 MW; 96CCD5A190C1FA5D CRC64;

Query Match 14.9%; Score 60; DB 2; Length 401;
Best Local Similarity 100.0%; Pred.No. 1.8e-49;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 IEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTTAPPHIPIMIAPAMHE 139
DB |||||||
81 IEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTTAPPHIPIMIAPAMHE 140

RESULT 4
ID Q9X213 PRELIMINARY; PRT; 394 AA.
AC Q9X213;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE DNA/pantothenate metabolism flavoprotein.
GN OrderedLocusNames=TW1687;
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,
Heldberg J.F., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 398:323-329(1999).
DR EMBL; AE001809; AAD36754.1; -;
DR PIR; A72223; A72223.
DR HSP; Q9SWE5; IE20.
DR TIGR; TM1687; -;
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 43804 MW; EBCC811C151ECFCA CRC64;

Query Match 3.0%; Score 12; DB 2; Length 394;
Best Local Similarity 100.0%; Pred.No. 0.013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 RPTNASSGKMG 223
DB |||||||
200 RPTNASSGKMG 211

RESULT 5
ID Q83NK7 PRELIMINARY; PRT; 410 AA.
AC Q83NK7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative flavoprotein.
GN OrderedLocusNames=TW408;
OS Tropheryma whipplei (strain TW08/27) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Micrococcineae; Cellulomonadaceae; Tropheryma.

OX NCBI_TaxID=218496;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22495039; PubMed=12606174; DOI=10.1016/S0140-6736(03)12597-4;

RA Bentley S.D., Maiwald M., Murphy L.D., Pallen M.J., Yeats C.A.,

RA Dover L.G., Norbertcak H.T., Besra G.S., Quail M.A., Harris D.E.,

RA von Herbay A., Goble A., Rutter S., Squares R., Squares S.,

RA Barrall B.G., Parkhill J., Rellman D.A.,

RT "Sequencing and analysis of the genome of the Whipple's disease

RT bacterium Tropheryma whippelii";

RL Lancet 361:637-644 (2003);

DR EMBL; BX251411; CAD67079.1; -;

DR HSSP; QSW55; IE20.

DR InterPro; IPR005252; Cons hypoth521.

DR InterPro; IPR007085; Dfp_C.

DR InterPro; IPR003382; Flavoprotein.

DR Pfam; PF04127; Dfp; 1.

DR Pfam; PF02441; Flavoprotein; 1.

DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.

KW Complete proteome.

SQ SEQUENCE 410 AA; 44783 MW; EC7BAFB9F23DDB29 CRC64;

Query Match 3.0%; Score 12; DB 2; Length 410;

Best Local Similarity 100.0%; Pred. No. 0.013; 0; Indels 0; Gaps 0;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 DVVIMAAVSDF 288

DB 268 DVVIMAAVSDF 279

|||||

RESULT 6

Q83MW7 PRELIMINARY; PRT; 417 AA.

AC Q83MW7

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Flavoprotein.

GN Name=dpf; OrderedLocNames=TW1361;

OS Tropheryma whippelii (strain Twist) (Whipple's bacillus).

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Micrococcineae; Cellulomonadaceae; Tropheryma.

OX NCBI_TaxID=203267;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=Twist;

RX MEDLINE=22784088; PubMed=12902375;

RA Raoult D., Ogata H., Audic S., Robert C., Suhr K., Drancourt M.,

RA Claverie J.-M.,

RT "Tropheryma whippelii Twist: a human pathogenic Actinobacteria with a

RT reduced genome";

RL Genome Res. 13:1800-1809 (2003).

DR EMBL; AS016851; AAO44458.1; -;

DR HSSP; Q9RC23; 1P3Y.

DR InterPro; IPR005252; Cons hypoth521.

DR InterPro; IPR007085; Dfp_C.

DR InterPro; IPR003382; Flavoprotein.

DR Pfam; PF04127; Dfp; 1.

DR Pfam; PF02441; Flavoprotein; 1.

DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.

KW Complete proteome.

SQ SEQUENCE 417 AA; 45584 MW; E2C79EC9B860BC63 CRC64;

Query Match 3.0%; Score 12; DB 2; Length 417;

Best Local Similarity 100.0%; Pred. No. 0.013;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 DVVIMAAVSDF 288

DB 275 DVVIMAAVSDF 286

|||||

us-08-957-709a-19.oligo.rup

RESULT 7

Dfp_METJA

ID DFP_METJA

AC Q58323;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Coenzyme A biosynthesis bifunctional protein coaBC (DNA/pantothenate

DE metabolism flavoprotein) [Includes: Phosphopantothenoylecysteine

DE decarboxylase (EC 4.1.1.36) (PPDC) (CoaC); Phosphopantothenate--

DE cysteine ligase (EC 6.3.2.5) (Phosphopantothenoylecysteine synthase)

DE (PPC synthetase) (PPCS) (CoaB)].

GN Name=coaBC; OrderedLocNames=MJ0913;

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

OC Methanocaldococcaceae; Methanocaldococcus.

OX NCBI_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,

RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,

RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.W., Glodek A.,

RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

RA Klenk H.-P., Fraser C.M., Smith H.O., Wosse C.R., Venter J.C.,

RA "Complete genome sequence of the methanogenic archaeon, Methanococcus

RA jannaschii.";

RT Science 273:1058-1073 (1996).

CC -!- FUNCTION: Catalyzes two steps in the biosynthesis of coenzyme A.

CC In the first step cysteine is conjugated to 4'-phosphopantothenate

CC to form 4'-phosphopantothenoylecysteine, in the latter compound is

CC decarboxylated to form 4'-phosphopantothoine (By similarity).

CC -!- CATALYTIC ACTIVITY: N-[(R)-4'-phosphopantothenoylecysteine =

CC pantotheine 4'-phosphate + CO(2).

CC -!- CATALYTIC ACTIVITY: CTP + (R)-4'-phosphopantothenoylecysteine

CC = CMP + Ppi + N-[(R)-4'-phosphopantothenoylecysteine.

CC -!- COFACTOR: Binds 1 FMN per subunit (By similarity).

CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; second step.

CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; third step.

CC -!- SUBUNIT: Homododecamer, the coaB domains form homodimers.

CC -!- SIMILARITY: In the N-terminal section; belongs to the HFCD (homo-

CC oligomeric flavin containing Cys decarboxylase) superfamily.

CC -!- SIMILARITY: In the C-terminal section; belongs to the PPC

CC synthetase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U67535; AAB98918.1; -;

DR PIR; A64414; A64414.

DR HSSP; P30197; IG63.

DR TIGR; MJ0913; -;

DR InterPro; IPR011256; Bac reg effector.

DR InterPro; IPR005252; CoaB hypoth521.

DR InterPro; IPR007085; Dfp_C.

DR InterPro; IPR003382; Flavoprotein.

DR Pfam; PF04127; Dfp; 1.

DR Pfam; PF02441; Flavoprotein; 1.

DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.

KW Complete proteome; Flavoprotein; FMN; Hypothetical protein; Ligase;

KW Lyase.

FT CHAIN 1 195 Phosphopantothenoylecysteine

FT decarboxylase.

```

FT CHAIN      196 403      Phosphopantothenate--cysteine ligase.
FT ACT_SITE   87          By similarity.
FT SEQUENCE   403 AA; 45670 MW; B8851AA81A115B94 CRC64;
SQ
Query Match      2.7%; Score 11; DB 1; Length 403;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALAE 228
Db 219 SSGKMGVALAE 229
|||||
RESULT 8
Q971S3 ID Q971S3 PRELIMINARY; PRT; 86 AA.
AC Q971S3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein STS144.
GN ORFNames=STS144;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / ?;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AF000985; BAB66347.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 86 AA; 10227 MW; CB80DBC0FDF13B93 CRC64;

Query Match      2.5%; Score 10; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 IERLKKLGVE 160
Db 49 IERLKKLGVE 58
|||||
RESULT 9
Q9JT98 ID Q9JT98 PRELIMINARY; PRT; 394 AA.
AC Q9JT98;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein NMA1916.
GN OrderedLocusNames=NMA1916;
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919; DOI=10.1038/35006655;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.M.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K.L., Quail M.A.,

```

```

RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162757; CAB85137.1; -.
DR PIR; D81819; D81819.
DR HSP; Q9SWE5; I20.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaABC_dfp; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 394 AA; 42238 MW; E19917C24E62B087 CRC64;

Query Match      2.5%; Score 10; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALA 227
Db 208 SSGKMGVALA 217
|||||
RESULT 10
Q9JYB7 ID Q9JYB7 PRELIMINARY; PRT; 394 AA.
AC Q9JYB7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE DNA/pantochenate metabolism flavoprotein.
GN OrderedLocusNames=NMB1658;
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307; DOI=10.1126/science.287.5459.1809;
RA Tettelin H., Saunders N.J., Heidelberg J.F., Jeffries A.C.,
RA Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F.,
RA Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R.T., Peterson J.D.,
RA Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D.,
RA Dougherty B.A., Mason T.M., Ciecko A., Parksey D.S., Blair E.,
RA Cittone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H.M.,
RA Qin H., Vamathevan J.J., Gill J., Scarlato V., Maignani V., Pizza M.,
RA Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R.,
RA Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002516; AAF42007.1; -.
DR PIR; A81058; A81058.
DR HSP; Q9SWE5; I20.
DR TIGR; NMB1658; -.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaABC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 42212 MW; B427D1901932AB91 CRC64;

Query Match      2.5%; Score 10; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALA 227

```



```
Db 208 SSGKMGVALA 217
|||||
Q9RQH7 PRELIMINARY; PRT; 399 AA.
AC Q9RQH7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Pantothenate metabolism flavoprotein homolog.
GN Name=dfp;
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RC STRAIN=ECD;
RX MEDLINE=20208566; PubMed=10746777;
RA Milohanic E., Pron B., Berche P., Gaillard J.L.;
RT "Identification of new loci involved in adhesion of Listeria
RL Microcytogenes to eukaryotic cells.";
DR EMBL; AF104226; AAF04763.1; -.
DR HSSP; Q9SWES; 1E20.
DR InterPro; IPR005252; Cons hypo521.
DR InterPro; IPR007085; Dfp_C.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
SQ SEQUENCE 399 AA; 4335 MW; 566E98FF6F07F7EB CRC64;

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVVTAGATRE 206
Db 190 VLVVTAGATRE 199
|||||

RESULT 13
Q92AI3 PRELIMINARY; PRT; 399 AA.
AC Q92AI3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Lin1939 protein.
GN OrderedLocusNames=lin1939;
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RC STRAIN=FROM N.A.
RX STRAIN=CLIP 11262 / Serovar 6a;
RC MEDLINE=21537279; PubMed=11679669; DOI=10.1126/science.1063447;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Knef J., Kuhn M., Kunst F., Kurapkat G.,
RA Madsen E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Norddick G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596170; CAC97169.1; -.
DR FIR; A11674; A11674.
DR HSSP; Q9SWES; 1E20.
DR ListList; LIN1939; -.
DR InterPro; IPR005252; Cons hypo521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 399 AA; 43475 MW; 2FEAD8EF3958B1F5 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVVTAGATRE 206
Db 190 VLVVTAGATRE 199
|||||

RESULT 14
Q71YJ1 PRELIMINARY; PRT; 399 AA.
AC Q71YJ1;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591981; CAC99903.1; -.

```

```
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Phosphophanthoenoylcysteine decarboxylase/phosphopantothenate--
DE cysteine ligase.
GN Name=coaBC; OrderedLocusNames=LMOF2365_1853;
OS Listeria monocytogenes (serotype 4b / strain F2365).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=263669;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15115801; DOI=10.1093/nar/gkh562;
RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,
RA Kolonay J.P., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,
RA Peterson J.D., White O., Nelson W.C., Nierman W.C., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,
RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,
RA Forberger H.A., Tran B., Kathariou S., Wonderling L.D., Uhlich G.A.,
RA Bayles D.O., Luchansky J.B., Fraser C.M.;
RT "Whole genome comparisons of serotype 4b and 1/2a strains of the food-
RT borne pathogen Listeria monocytogenes reveal new insights into the
RT core genome components of this species.";
RL Nucleic Acids Res. 32:2386-2395(2004).
DR EMBL; AE017328; AAT04623.1; -.
DR GO; GO:0016874; F.ligase activity; IEA.
DR InterPro; IPR005252; Cons hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome; Ligase.
SQ SEQUENCE 399 AA; 43490 MW; 1EF88D4C8C2A8344 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVTAGATRE 206
Db |||||
190 VLVTAGATRE 199

RESULT 15
Q976C4 PRELIMINARY; PRT; 412 AA.
AC Q976C4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein ST0257.
GN OrderedLocusNames=ST0257;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AF000982; BAB65223.1; -.
DR InterPro; IPR005252; Cons hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
```

```
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 412 AA; 45517 MW; EB28B9F11338947 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 PATANTISKI 109
Db |||||
98 PATANTISKI 107

Search completed: July 2, 2005, 12:38:11
Job time : 82.0233 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:17:13 ; Search time 23.0698 Seconds
(without alignments)
1680.787 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 1998
Sequence: 1 MLHHVLIYATKSRKLVGKK.....MKKRELAERIWDIEIKXLS 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1667.5	83.5	401	2 G71018	pantothenate metab
2	1649.5	82.6	401	2 E75114	dna/pantothenate m
3	735	36.8	404	2 D69455	pantothenate metab
4	731	36.6	386	2 D69029	pantothenate metab
5	648.5	32.5	413	2 B90162	DNA/pantothenate m
6	627.5	31.4	403	2 A64414	pantothenate metab
7	626	31.3	392	2 C84215	pantothenate metab
8	605	30.3	398	2 C97112	flavoprotein invol
9	543.5	27.2	390	2 C70201	pantothenate metab
10	535.5	26.8	388	2 B70371	pantothenate metab
11	535	26.8	437	2 A72498	probable DNA/panto
12	506	25.3	399	2 A11674	pantothenate metab
13	494.5	24.7	404	2 F83963	flavoprotein dfp l
14	494	24.7	399	2 A11302	pantothenate metab
15	490.5	24.5	399	2 A82351	DNA/pantothenate m
16	483.5	24.2	402	2 B82982	DNA/pantothenate m
17	482.5	24.1	394	2 A72223	pantothenate metab
18	473.5	23.7	406	2 D69878	pantothenate metab
19	473	23.7	394	2 A81058	DNA/pantothenate m
20	464	23.2	394	2 D81819	hypothetical prote
21	459.5	23.0	400	2 G64104	pantothenate metab
22	457	22.9	405	2 A10006	DNA/pantothenate m
23	453.5	22.7	399	2 F89893	hypothetical prote
24	449	22.5	404	2 A13538	phosphopantothen
25	447	22.4	422	2 H75501	DNA/pantothenate m
26	439	22.0	402	2 S75082	pantothenate metab
27	437.5	21.9	401	2 AD2615	pantthotenate metab
28	435.5	21.8	410	2 AD2194	pantothenate metab
29	430.5	21.5	418	2 E70899	pantothenate metab

30	424.5	21.2	430	1 A65165	pantothenate metab
31	424.5	21.2	430	2 C86040	pantothenate metab
32	424.5	21.2	430	2 B91193	pantothenate metab
33	420.5	21.0	407	2 AC0971	conserved hypochet
34	418.5	20.9	419	2 G86976	probable flavoprot
35	408	20.4	367	2 C97397	pantothenate metab
36	404.5	20.2	409	2 H82842	DNA/pantothenate m
37	397	19.9	412	2 F87709	DNA/pantothenate m
38	351	17.6	425	2 E71889	pantothenate metab
39	350	17.5	425	2 A64625	pantothenate metab
40	301	15.1	384	2 F81354	DNA /pantothenate
41	260	13.0	183	2 H95142	flavoprotein [impo
42	260	13.0	183	2 F98010	N-terminal region
43	239	12.0	234	2 E98010	Dfp protein homolo
44	238	11.9	229	2 G95142	conserved hypochet
45	224.5	11.2	231	2 E86696	flavoprotein [impo

ALIGNMENTS

RESULT 1

G71018
pantothenate metabolism flavoprotein dfp homolog PH1444 - Pyrococcus horikoshii
N;Alternate names: probable aspartate 1-decarboxylase activase
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: G71018
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: G71018
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-401 <RAW>
A;Cross-references: UNIPROT:O59114; GB:AP000006; MID:g3236133; PIDN:BAA30551.1; PID:g32
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1444

Query Match	83.5%;	Score	1667.5;	DB 2;	Length	401;			
Best Local Similarity	83.4%;	Pred. No.	3.4e-92;						
Matches	337;	Conservative	33;	Mismatches	29;	Indels	5;	Gaps	4;
Qy	1	MLHHVKLIYATKSRKLVGKKIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKII	59						
Db	1	MLHHVRIYAKKGRKLVGKKIVLAI PGSIAAVECVKLARELIRHGAEVHVMTPSATKII	60						
Qy	60	HPYAMNLPNGNPVITEITGPIEHVELAGEHENKADLILVCPATANTISKIACGIDDDTPVT	119						
Db	61	HPYAMEFATGNPVITEITGPIEHVELAGEHENKADLILVCPATANTISKIACGIDDDTPVT	120						
Qy	120	TVVTTAPPHIPMIAPAMHETMYRHPFIVRENIERLKKLGVEFTGPRIEBEGRAKVASIDEI	179						
Db	121	TVVTTAPPHIPMIAPAMHESMYKHPIVRENIERLKKLGVEFTGPRIEBEGRAKVASIDEI	180						
Qy	180	VYRVIKLHKHKTLEGKRLVATAGATREYIDPIRFTNASSGKMGVALAEADPRGA-VTL	238						
Db	181	VYRVIRKLPKTLGKRLVATAGATREYIDPIRFTNASSGKMGVALAEAEPRGAEVTL	240						
Qy	239	IRTKGSVKAFPRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDPFRPKIKAEKGI	298						
Db	241	IKTKGSVNSFVENQI--QVETVEEMLAAIEKELTEKKYDVVIMAAVSDPFRPKIKAEKDI	298						
Qy	299	KSGRSITIELVPXNPKIIDRIEIQPNVFLVGKAEATSKEKLIEBEGKQIERAKADLVVG	358						
Db	299	KSNKSITIELVP-NPKIIDRIEKLQDPVFLVGKAEATTMEKLIQEAKKQIERAKSDIVIG	357						
Qy	359	NTLEAFSGSEENVVLIGRDFTKELPKMKRELAERIWDIEIKXL	402						

QY 138 HETMYRHPVIRENLERKLVGVEFGIPRIEGRKAVASIDEIVYRVVKKLHKKTLEGRV 197
DB 117 HEAMY--AAAENIMLKKEGVIFIEPRMDEGKAKFPDITIVLEAMROTSSORLOKRV 174
QY 198 LVTAGATREYIDPIRITINASSGKMGVALAEADFRGA-VTLIRTKGSV-----KAFRI 250
DB 175 LVSLGGTYEPIDPVGITNRSSGKMGLAIRRAYIEGADVTVVAGTVSVIEIPQLRSFR- 233
QY 251 RKTIKLVETVEEMLSAIENELRSKKYDVIMAAVSDFRPKIKAEKIKSGRSITIELVP 310
DB 234 -----AETAEEAERVELVAD--HDVFTSAAAVADFKP-VYTERKISSSEFSVELRP 284
QY 311 XNPKIIDRIKEIOPNVFLVGFKA--TSKEKLEEGKROIERAKADLVVGN--TLAEGS 366
DB 285 -NPKVIGIAREINPEAFIVGFKAEDVDNEALVESARKQIRESGVDMVAVDSVEGFGS 343
QY 367 EENQVVLIGRDFKELPKMKRELAERIMDEI 398
DB 344 DRNRALIVS-DMVTEPLMEKEELASIIIDEV 374
RESULT 5
B90162
DNA/pantothenate metabolism flavoprotein (dfp) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: B90162
R:Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, B.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: B90162
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-413 <KUR>
A:Cross-references: UNIPROT:Q98050; GB:AE006641; NID:g13813345; PIDN:AAK40553.1; GSPDB:G
C:Genetics:
A:Gene: dfp
C:Superfamily: pantothenate metabolism flavoprotein dfp
Query Match 32.5%; Score 648.5; DB 2; Length 413;
Best Local Similarity 39.7%; Pred. No. 1.7e-31;
Matches 166; Conservative 81; Mismatches 144; Indels 27; Gaps 11;
QY 1 MLHHVKLIYATKSRKLVGKIVXXPGSTA--ALDVKACEGLIRHGAHVHVMSEAATK 57
DB 5 MTHPSKKIIGSISNELADKKILLAVGTGSVAIYKSLDL--ARSLMRNGAEVSVIISKDAK 62
QY 58 IHPYAWNLPNGPNVITEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTTP 117
DB 63 LISPENFKWATGNVVTKLTDLEHVSLEADN---DVMIVAPSTANTWVKIAYGIADTP 118
QY 118 VTTVTTAFPHI-----PIMAPAMHETMYRHPVIRENLERKLVGVEFGIPRIEGRKAV 173
DB 119 IT---ATALNFVGKKPLIIVPSMHLQWYISPOVADAADRLKRGIVEVIEPEIVGDLAHY 175
QY 174 ASIDEIVYRVIK-KLHKKTLEGRKVLVTAGATREYIDPIRITINASSGKMGVALAEADP 232
DB 176 PKLEYITSRITSVYLRGKDLSGFNILATAGPTREYIDSVRFINPSSGTMGISIANEAYF 235
QY 233 RGA-VTLIRTKGSVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVIMAAVSDFRPK 291
DB 236 RGAKVRVICGPTSSKLEPYVKDVIYVETTEEMLNEVVKSIENGKPNVILAGAPADYKFK 295
QY 292 IKAEGKIKSGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAET--SKEKLIBEGKROIE 349
DB 296 NKSDTKIDSHTEIPKVELERTPKISEYIRKY--NVLVGSFASVTNSDEELIEKAKIKMR 353
QY 350 RAKADLVVGNLLE-----ARGSENOVVLIGRD-FTKELPKMKRELAERIWDEIKL 402
DB 354 RHGFDLIVANNVRRKIDGFSSEYNEVIVIDKNGDVRKIPKNFKTVIARKILDIVKSQL 411

RESULT 6
A64414

pantothenate metabolism flavoprotein MJ0913 - Methanococcus jannaschii
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: A64414
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A.; Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:868087
A:Accession: A64414
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-403 <BUL>
A:Cross-references: UNIPROT:Q59323; GB:U67535; GB:L77117; NID:g2826348; PIDN:AAB989918.1
C:Genetics:
A:Map position: REV845792-844581
C:Superfamily: pantothenate metabolism flavoprotein dfp
C:Keywords: flavoprotein

Query Match 31.4%; Score 627.5; DB 2; Length 403;
Best Local Similarity 39.6%; Pred. No. 3e-30;
Matches 163; Conservative 82; Mismatches 138; Indels 29; Gaps 16;

QY 1 MLHHVKLIYATKSRKLVGKIVXXPGSTAALDV-KACEGLIRHGAHVHVMSEAATKII 59
DB 5 IMHPTKLLGKTKSKLLENKKILVAVTSSIAAETPKLMRELIRHGAHVYCIITEETKKII 64
QY 60 HPYAWNLPNGPNVITEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVT 119
DB 65 GKELAFGCGNEVEEITGDIHIL-----YNECDCLLIYPATANIISKINLGADNVN 120
QY 120 TVVTTAFPHIPIAMAPAMHETMYRHPVIRENLERK-KLVGVEFGIPRIEGRKAVSIDE 178
DB 121 TTALMFFGNGKPIFIVPAMHENMFN--AIKRHIDKLEKDKIYIISPKEFGKAKVANIED 178
QY 179 IVYRVVTKL-HKKTLEGRKVLVTAGATREYIDPIRITINASSGKMGVALAEADFRG-AV 236
DB 179 VYKAVTEKIGNLKKEGNRVLILNGGTVEFDKRVVISNLSGKMGVALAEAFCKEGFYV 238
QY 237 TLIRTKGSVKAFRIRKIKLVETVEEMLS-AIENELRSKKYDVIMAAVSDFRPKIAE 295
DB 239 EVITAMGLEPPYYIK--NHKVLTAKEMLNKAIB---LAKDFDIIISAAISDTVE-SFE 292
QY 296 GKIKSGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKA--TSKEKLIBEGKROIERAKA 353
DB 293 GKLSSEELILKL-KENPKVLELRIRYKDKVIIGFKAENYLNDEKELINRAKERLNKYNL 351
QY 354 DLVVGNTLEA--FGSENOVVLIGROFTK-ELPDM--KRELAERIWDEIEK 400
DB 352 NMIIANDLSKHVGGDYIEVYII---TKYEVEKISGSKEISERIVEKVK 399

RESULT 7
C84215

pantothenate metabolism flavoprotein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: C84215
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Laasy,
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabb,
Jung, K.H.; Alan, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L
A.; Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: C84215

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-392 <STO>
A:Cross-references: UNIPROT:Q9HRS1; GB:AE004437; NID:g10580167; PIDN:AAG19087.1; GSPDB:C
C:Genetics: dfg
A:Gene: dfg
C:Superfamily: pantothenate metabolism flavoprotein dfg

Query Match 31.3%; Score 626; DB 2; Length 392;
Best Local Similarity 38.1%; Pred. No. 3.5e-30;
Matches 151; Conservative 73; Mismatches 146; Indels 26; Gaps 11;

QY 16 LVGKIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAATKIHPYAWNLPTGNPVIT 74
DB 2 LSGVNVAVGVTSIAAKVVEFVHELRGACVRAVMTESAGIHPHWAPEFATENPVVT 61

QY 75 EITGPIEVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTTAF-PIHIPMI 133
DB 62 EITGVPHVELCG-RDGWADVFPVAPATANTVGKIAAAVDDSPVITCVTTAVGADVPPVV 120

QY 134 APAMHETMYRHPVIRENTERLKKLVGFIPIRIGRAKVASIDIVYRVIKLHKHTLE 193
DB 121 VPAMHEPMYDHGVRDAIDRVSSGVSVVDDPRIIEGKAKLPRESTIVHETARAAGEQPLA 180

QY 194 GKRVLVTAGATREYIDPIRFITNASSGKMGVALAEAEADFRGA-VTLIRTK---GSVKAFR 249
DB 181 GTHVVVTSGATSEADIPVRVLNTRASGRTRGRAVAAACYVRGARVTLVHDASDGGAVPYAD 240

QY 250 IRKIKLVETVBEMLSAENELRSKKYDVVIMAAVSDFRPKIKAEKIKSGR-SITIEL 308
DB 241 VRDVSSAAEMTAATLTACADA-----DALVSAAISDYTVE-AADEKLRSREDVALSL 293

QY 309 VPXNPKIIDRIKEIQNPVFLVGFKAET-----SKEKLIBEGKQIERAKADLVVGNL 361
DB 294 EPTR-KLVGADVDDNPDLPIVGFKAETPADGDTAGDSDSMVAARSLLQRLGLAFVAVDA 352

QY 362 EAFGSEENQVLIGRDTFKELPKMKK---RELAERI 394
DB 353 GVMGNDETALFVTDVSDVSEYAGHKRGLGARIAERL 388

RESULT 8
C97112
Flavoprotein involved in pantothenate metabolism, YLOI B. subtilis ortholog [imported]
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: C97112
R.Nolling, J.; Brennet, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
; Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: C97112
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <KUR>
A:Cross-references: UNIPROT:Q97IC8; GB:AE001437; PIDN:AAK79686.1; PID:g15024686; GSPDB:C
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1720
C:Superfamily: pantothenate metabolism flavoprotein dfg

Query Match 30.3%; Score 605; DB 2; Length 398;
Best Local Similarity 39.8%; Pred. No. 6.4e-29;
Matches 165; Conservative 76; Mismatches 124; Indels 50; Gaps 17;

QY 19 KIVVXXPGSIA--ALDV-----KACEGLIRHGAHVHVMSEAATKIHPYAWNLPTGN 70
DB 4 KNVIVGSGVIAVYKALDVISRLKADFG-----VDVIMTNSASKFTVPLSFQSLGN 56

QY 71 PVITEI-----TGFEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTTA 125
DB 57 KVINDMFEPEKWEIQHISLA---KKAAILVVPATANIIGKVANGIADMLSTTMAA 112

QY 126 PPHIPIMLAPAMHETMYRHPVIRENTERLKKLVGFIPIRIGRAKVASI-----DE 178
DB 113 YGRCPVPAPAMTMYNPNVQNNIKKLDGYGIFIEP--EKGLACGDVGEGLADTE 170

QY 179 IVYRVIKKL--HKKTEGKRVLTAGATREYIDPIRFITNASSGKMGVALAEAEADFRGA- 235
DB 171 LIYENIKSLLYNKRDLAGKQVWVTAGPTIAPIDPVRFTTNHSTGKMGVAIAEARDGAE 230

QY 236 VTLIRTKGSVK-APRIIRKIKLVETVBEMLSAENELRSKKYDVVIMAAVSDFRPKIKA 294
DB 231 VTLISGETSLKPPFGVDFI--KVTNSSEMKEVLEKFGSG--DIVIKSAVADYKAKNYS 286

QY 295 EGKI-KSGRSITIELVPXNPKIIDRIKEIQNPVFLVGFKAETSKEKLIBEGKQIERAKA 353
DB 287 ELKIKASDNLNIIEFVKDN-DILKKGLEIKKHQILVGFPAESN--DLIENAVGKLRKNL 343

QY 354 DLVVGNTL-----EAFGSEENQVLIGRDTFK-ELPKMKKRELAERINDBIEKXLS 403
DB 344 DIVVANDILSKDAGFASDENRVILGSDGSKLELDKMSKRKVAENLFDLLSKRS 398

RESULT 9
C70201
pantothenate metabolism flavoprotein dfg homolog BB0812 - Lyme disease spirochete
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: C70201
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whit
son, D.; Peterson, J.; Korlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugl,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horat, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: C70201
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-390 <LSE>
A:Cross-references: UNIPROT:O51752; GB:AE001179; GB:AE000783; NID:g2688738; PIDN:AAAC6714
A:Experimental source: strain B31
C:Superfamily: pantothenate metabolism flavoprotein dfg

Query Match 27.2%; Score 543.5; DB 2; Length 390;
Best Local Similarity 36.3%; Pred. No. 2.9e-25;
Matches 144; Conservative 81; Mismatches 135; Indels 37; Gaps 14;

QY 19 KIVVXXPGSIAAL-DVKACEGLIRHGAHVHVMSEAATKIHPYAWNLPTGNPVITEIT 77
DB 5 KHLIGICGGIASYKSVIVSVSLVGLYKVKVIMTQNAKTFITPLTLETISKNIITNLW 64

QY 78 GF-----IEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTTTAPPHIPMI 133
DB 65 DLDHNEVEHIKIA---KWAHLILVIPATYNTISKIAGIADALTITIAS--TAPTFF 118

QY 134 APAMHETMYRHPVIRENTERLKKLVGFIPIRIGRIGRIGR-----RAKVASIDIVYRVIK 185
DB 119 AIFAMNIMYSPILKENIKKLTNYKFTIEP--DKGLACSSNALGRKNKEDIKIILN 176

QY 186 KLHKKT-LEGKRVLTAGATREYIDPIRFITNASSGKMGVALAEAEADFRGA-VTLI---R 240
DB 177 EPNQKDYLNKXKILITASRTELEIDPIRVFSNTSTGKMGFCLAQEAQVAVKLAQVTLITGPT 236

QY 241 TKGSVKAFRIKIKLVETVBEMLSAENELRSKKYDVVIMAAVSDFRPKIKAEKIKS 300
DB 237 NENDPEGVNIIKIKTAMEMYKALKIY-----NKFETIIGAAVADFPKGFHFNISKIK 290

QY 301 GR-SITIELVPXNPKIIDRIKEIQ-PNVFLVGFKAETSKEKLIBEGKQIERAKADLVV 357
DB 291 NKNINLYKLW-KNPDIIOHGNHKNQIQVIGFCAENSK-NLIQAKAKLKKKNDLFII 348

QY 358 GNTLEAFGSEENQVLIGRDTFKELPKMKKRELAERI 394


```
Db      63 PDEKSSVVAHIDLA-----DWADLVIVAPATANLKGKMGANGIADDMWVTTILAT--EAPV 116
Qy      132 MIAPAMHETMYRHPVIRENIERLKKLVGEFIPRIEE-----GRAKVASIDEIVRV 183
Db      117 WVAPAMNVHMIQHPAVIRINRLYADGVRFIEP--EEGYLACGYVGRGLEPEKIVLRI 174
Qy      184 IK--KLHKKTLEGGKVLVTAGATREVIDPIRITNASSGKMGVALAEADFRGA-VTLIR 240
Db      175 AEFFQEDKLLQGGKVLVTAGATREKLDPVRIFTNHSITGKMGFSAESAAHGANVTLIT 234
Qy      241 TKGS-----VKAFRIRIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKA 294
Db      235 TSKTLVPVPGVEA-----VIVESADEMYQAVLE--RKAEQHIFWTAADVYTPANVS 285
Qy      295 EGKI-KSGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKEKLIEGKRQIERAKA 353
Db      286 EQKIKKQPGDFTIEMKRTKDILLEGQNKQTADQVVGFAAET--ENLETNALKCLTSKNA 343
Qy      354 DLVVGNTLE---AFGSEENQVVLIGRDTKE-LPKMKKRELAERIWDIEKKL 402
Db      344 DMIVANNISEAGAGSGDTNIVTYRKGSGNEALPLDKKEVAEHIIIEAANFL 397

RESULT 13
F83963
flavoprotein dfp [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: F83963
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F83963
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-404 <STO>
A:Cross-references: UNIPROT:Q9K9Y4; GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA8062
A:Experimental source: strain C-125
C:Genetics:
A:Gene: dfp
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match      24.7%; Score 494.5; DB 2; Length 404;
Best Local Similarity 33.5%; Pred. No. 2.5e-22;
Matches 138; Conservative 84; Mismatches 149; Indels 41; Gaps 15;

Qy      16 LVGKTIVXXPGSIAALDVKA-CEGLIRHGAEVHVMSEAATKIHPYAWNLPNGPVIT 74
Db      2  LQGRVVLGVSGGIAAFKSAAFASKLVQAGAEVQVVMTEGAKKFTVPLTFQALTRHPYVD 61

Qy      75 EI-----TGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDTPTVTVVTTAFPHI 129
Db      62 DTFSEPDPEIAHIQLA----DWADVITATPATANLIGLANGVADDMLSMTLLAT--KA 115

Qy      130 PMIAPAMHETMYRHPVIRENIERLKKLVGEFIPRIEE-----GRAKVASIDEIVRV 183
Db      116 PIVLAPAMNVNVEHPAVQRNQMLAKDGYRLLEPGAGYLACGWIGKGMPEPEDILKTI 175

Qy      184 IKKLH-----KKTLEGRVLVTAGATREYIDPIRITNASSGKMGVALAEAE--DFRGAVTL 238
Db      176 --EVMFTPPSSLAGKKVITAGTQETIDPIRFTNRSRSGKMGVALTTAAKADFQGNVTL 233

Qy      239 IRTKGSV-KAFRIRIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGK 297
Db      234 ISGPTSLERPDGVSVMVK--SAQDMYEAFLAEFSDA--DVWIKTAADVYPRVVEHQK 289

Qy      298 IKSRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKEKLIEGKRQIERAKADLVV 357
Db      290 VKKKGDWIELERTVDILKTGEEKESQFLVGFAAES--QEVETVQAQKLEKKNADMIV 347

Qy      358 GNTL----EAFGSEENQV-VLIGRDTFKELPKMKKRELAER----IWDIEIK 400
```

```
Db      348 ANNVTEGAGQTDTNRVTYVFQGDVKKPLMTKDEVAHRLIMMISEQLEK 399

RESULT 14
A11302
pantothenate metabolism flavoprotein homolog lmo1825 [imported] - Listeria monoc
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: A11302
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Me
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11302
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-399 <GLA>
A:Cross-references: UNIPROT:Q8Y674; GB:NC_003210; PIDN:CAC99903.1; PID:g16411279; GSPDB:
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo1825
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match      24.7%; Score 494; DB 2; Length 399;
Best Local Similarity 34.8%; Pred. No. 2.6e-22;
Matches 144; Conservative 70; Mismatches 152; Indels 48; Gaps 16;

Qy      18 GKKIVXXPGSIAALDVKACBG-LIRHGAEVHVMSEAATKIHPYAWNLPNGPVITEI 76
Db      3  GKNILLAVSGGIADVYKVALTSKLTQAGANVKVMVMTAAHQAEFVPLSPQVLSKNDVYTT 62

Qy      77 -----TGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDTPTVTVVTTAFPHI 131
Db      63 PDEKSSVVAHIDLA----DWADLVIVAPATANVIGKMGANGIADDMWVTTILAT--EAPV 116

Qy      132 MIAPAMHETMYRHPVIRENIERLKKLVGEFIPRIEE-----GRAKVASIDEIVRV 183
Db      117 WVAPAMNVHMIQHPAVIRINRLYADGVRFIEP--EEGYLACGYVGRGLEPEKIVLRI 174

Qy      184 IK--KLHKKTLEGGKVLVTAGATREYIDPIRITNASSGKMGVALAEADFRGA-VTLIR 240
Db      175 AEFFQEDKLLQGGKVLVTAGATREKLDPVRIFTNHSITGKMGFSAESAAHGANVTLIT 234

Qy      241 TKGS-----VKAFRIRIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKA 294
Db      235 TSKALVPVPGVEA-----IIVESAEEMHQAV-NE-RKVSQDIFWMTAAVADYTPAQVS 285

Qy      295 EGKI-KSGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKEKLIEGKRQIERAKA 353
Db      286 DQKIKKQPGDFTIAMKRTKDILLEGQNKQTSEQVVGFAAET--ENVEANARKCLTSKNA 343

Qy      354 DLVVGNTLE---AFGSEENQVVLIGRDTKE-LPKMKKRELAERIWDIEKKL 402
Db      344 DMIVANNISEAGAGSGDTNIVTYRKGSGSEALPILDKKEVAEHIIIEAANFL 397

RESULT 15
A82351
DNA/pantothenate metabolism flavoprotein VC0215 [imported] - Vibrio cholerae (strain NI
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: A82351
R:Heidelberg, J.F.; Easen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
```


Search completed: July 2, 2005, 12:25:15
Job time : 25.0698 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:17:13 ; Search time 8.93023 Seconds
(without alignments)
1680.787 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDKIRKELIIEPFSEEE.....PYRGNVQSGSTRLAFSKRKKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	768	95.3	156	E71216	dCTP deaminase (EC
2	703	87.2	154	G75030	dCTP deaminase (EC
3	243	30.1	180	A70439	probable dCTP deam
4	242.5	30.1	177	H81695	deoxycytidine triph
5	242	30.0	173	S28382	probable dCTP deam
6	232.5	28.8	181	D72724	probable dCTP deam
7	216	26.8	191	T36613	deoxycytidine triph
8	207.5	25.7	206	A90160	dCTP deaminase (EC
9	199	24.7	200	A69114	dCTP deaminase (EC
10	198	24.6	190	B70526	dCTP deaminase (EC
11	197.5	24.5	190	H87222	probable deoxycyti
12	189.5	23.5	186	A81272	probable dCTP deam
13	184.5	22.9	195	G84184	deoxycytidine triph
14	181	22.5	190	D64566	dCTP deaminase (EC
15	180	22.3	188	E71860	dCTP deaminase (EC
16	177	22.0	193	A10185	dCTP deaminase (EC
17	174	21.6	206	C84942	dCTP deaminase (EC
18	172	21.3	188	A83210	probable deoxycyti
19	167.5	20.8	193	AF0770	dCTP deaminase (EC
20	165	20.5	193	H90987	2'-deoxycytidine 5
21	165	20.5	193	C85833	2'-deoxycytidine 5
22	164.5	20.4	191	G82765	deoxycytidine triph
23	163	20.2	193	A42940	dCTP deaminase (EC
24	159	19.7	163	H72759	probable dCTP deam
25	155	19.2	188	D81149	deoxycytidine triph
26	154	19.1	204	F64353	dCTP deaminase (EC
27	153	19.0	150	D69081	deoxycytidine 5-tri
28	150.5	18.7	190	D81717	deoxycytidine triph
29	149.5	18.5	190	B71565	probable dCTP deam

30	149.5	18.5	195	1	A64050	dCTP deaminase (EC
31	148	18.4	190	2	C97712	hypothetical prote
32	145	18.0	188	2	E71715	probable dCTP deam
33	138	17.1	190	2	F72084	dCTP deaminase (EC
34	138	17.1	190	2	F86539	dCTP deaminase (im
35	137.5	17.1	199	2	AB1815	dCTP deaminase (im
36	137	17.0	172	2	T44356	probable dCTP deam
37	135.5	16.8	193	1	S75588	dCTP deaminase (EC
38	134	16.6	173	2	A96903	deoxycytidine triph
39	128.5	15.9	157	2	C97049	deoxycytidine 5-tri
40	127.5	15.8	166	2	A99475	deoxycytidine triph
41	126	15.6	168	2	C69388	probable dCTP deam
42	125.5	15.6	1145	1	GNLJEV	HIV-1 retropepsin
43	125.5	15.6	1146	1	GNLJ22	HIV-1 retropepsin
44	125.5	15.6	1146	1	GNLJEW	HIV-1 retropepsin
45	121	15.0	147	2	D95002	hypothetical prote

ALIGNMENTS

RESULT 1

E71216
dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C:Accession: E71216
R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki,
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch:
DNA Res. 5, 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71216
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-156 <KAW>
A:Cross-references: UNIPROT:O57706; GB:AP0000007; NID:g3236134; PIDN:BAA31124.1; PID:g32
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1997
C:Keywords: hydrolase

Query Match	95.3%	Score	768	DB	2	Length	156
Best Local Similarity	93.6%	Pred. No.	3.6e-65				
Matches	146	Conservative	4	Mismatches	6	Indels	0
Gaps	0						
Qy	1	MLLPDKIRKELIIEPFSEESLQAGYDLRVGREAFVKGKLDVEKEGKVPPREYALI	60				
Db	1	MLLPDKIRKELIIEPFSEESLQAGYDLRVGREAFVKGKLDVEKEGKVPPREYALI	60				
Qy	61	LTLEIRKLPDDVGMKIRSSLAEGVIGSFAWVDPGWDGNLTMLYNASNEPVELRYGE	120				
Db	61	LTLEIRKLPDDVGMKIRSSLAEGVIGSFAWVDPGWDGNLTMLYNASNEPVELRYGE	120				
Qy	121	RFVQIAFIRLEGPARNPNRYGQSTRLAFSKRKKL	156				
Db	121	RFVQIVFIRLEDPARNPNRYGQSTRLAFSKRKKL	156				

RESULT 2

G75030
dCTP deaminase (EC 3.5.4.13) PAB1164 [similarity] - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: G75030
R:Anonymous, Genoscope
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A:Reference number: A75001
A:Accession: G75030
A>Status: preliminary
A:Molecule type: DNA

A;Residues: 1-154 <KAW>
A;Cross-references: UNIPROT:Q9UXS8; GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB5068
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-177 <STO>
A;Cross-references: UNIPROT:Q9KFV3; GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BAB040
A;Experimental source: strain C-125
C;Genetics:
A;Gene: dcd; PAB1164
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 87.2%; Score 703; DB 2; Length 154;
Best Local Similarity 85.6%; Pred. No. 4.7e-59;
Matches 131; Conservative 13; Mismatches 59; Indels 0; Gaps 0;
Qy 1 MLLPDKIRKEILIPFSEESLOPAGYDLRVGREAFAVKGKIDIVEKEGKVIPPPEYALI 60
Db 1 MLLPDKIRKEILIPFSEESLOPAGYDLRVGKEAYIQKIDIVEKEGKVIPPPEYALI 60
Qy 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSAFVDPDGDGNTLMLYNASNEPVELRYGE 120
Db 61 LTLERIKLPDDINGDMKIRSSILAREGVLGSAFVDPDGDGNTLMLYNASEKEVILRYKE 120
Qy 121 RPFQIAFLRLEGPARNPYRGNYQGSSTRLAFSKR 153
Db 121 RPFQIAFLRLEAPAKNPYRGNYQGSRRIVLSKR 153

RESULT 3
A70439
probable dCTP deaminase (EC 3.5.4.13) dcd [similarity] - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
A;Accession: A70439
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ovi
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: A70439
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-180 <AQF>
A;Cross-references: UNIPROT:Q67539; GB:AE000747; NID:g2983944; PIDN:AAC07499.1; PID:g298
A;Experimental source: strain VF5
C;Genetics:
A;Gene: dcd
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 30.1%; Score 243; DB 2; Length 180;
Best Local Similarity 37.7%; Pred. No. 1.2e-15;
Matches 61; Conservative 30; Mismatches 53; Indels 18; Gaps 5;
Qy 1 MLLPDKIRK-----EILIPFSEESLOPAGYDLRVGRE-APVKGK-LIDVEKEGKV-- 51
Db 1 MILDSRILIEKELKVPVPSHVQSSLDLRLGNQIALYEGEVIDVKKGKGVRI 60
Qy 52 -----IPREYALILTLERIKLPDDVGMGDMKIRSSILAREGV-IGSAFVDPDGDGNTL 102
Db 61 LEFEYFDPMPKQFLATLTLEYISLPVYVTAPEGRSSILGRGLFIENAGWVDAGFEGQI 120
Qy 103 TMLYNASNEPVELRYGSRFVQIAFLRLEGPARNPYRGNYQG 144
Db 121 TLELFNANDRPIRLYRGHRIQCLVFAFLDRPPEVYSKYGK 162

RESULT 4
H83695
deoxycytidine triphosphate deaminase BH0368 [imported] - Bacillus halodurans (strain C-1
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
A;Accession: H83695
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: H83695
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-177 <STO>
A;Cross-references: UNIPROT:Q9KFV3; GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BAB040
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0368
C;Superfamily: dCTP deaminase

Query Match 30.1%; Score 242.5; DB 2; Length 177;
Best Local Similarity 37.2%; Pred. No. 1.4e-15;
Matches 58; Conservative 30; Mismatches 47; Indels 21; Gaps 6;
Qy 10 KEILIPFSEESLOPAGYDLRVGREAFV-----KGLIDIVEK-----EGKVVIPP 54
Db 15 KEILITLTERIQIPASVDLRGLPH-FVTIDDSKEAVISFERPIRYREWTTSDETVILPP 73
Qy 55 REYALILTLERIKLPDDVGMGDMKIRSSILAREGV-IGSAFVDPDGDGNTLMLYNASNEP 113
Db 74 HTFLATTMETVKLPNHLTAFVEGRSSVGRGLGFIONAGWVDGFGNGQITLLELFNANRLP 133
Qy 114 VELRYGSRFVQIAFLRLEGPARNPYRGNY---QGST 146
Db 134 IELPIGRRICQLVFAEVTGEVA-PYQGYLFOKCAT 168

RESULT 5
S26382
probable dCTP deaminase (EC 3.5.4.13) [similarity] - Desulfurolobus ambivalens
N;Alternate names: hypothetical protein 3 lig-region
C;Species: Desulfurolobus ambivalens
C;Date: 22-Nov-1993 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C;Accession: S26382
R;Kletzin, A.
Nucleic Acids Res. 20, 5389-5396, 1992
A;Title: Molecular characterisation of a DNA ligase gene of the extremely thermophilic a
A;Reference number: S26382; MUID:93065206; PMID:1437556
A;Accession: S26382
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-173 <KLE>
A;Cross-references: UNIPROT:Q02103; EMBL:X63438; NID:g40784; PIDN:CAA45033.1; PID:g40785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1991
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 30.0%; Score 242; DB 2; Length 173;
Best Local Similarity 35.9%; Pred. No. 1.5e-15;
Matches 56; Conservative 30; Mismatches 56; Indels 14; Gaps 3;
Qy 12 ILIPFSEESLOPAGYDLRVG-----REAFVKGK-----LIDVEKEGKVIPPPEYA 58
Db 17 IVISPLTQDTITRENGVDLRVGGELARFKKTDIYEDGKDRSFYEIEKGDFIIYPNSHV 76
Qy 59 LILTLERIKLPDDVGMGDMKIRSSILAREGVIGSAFVDPDGDGNTLMLYNASNEPVELRY 118
Db 77 LLVTEEVKLPNDVMAFVNLRSSIFARLGLFVPPTVDAGFEGQITIEVL-GSAFPVKIR 135
Qy 119 GERFVQIAFLRLEGPARNPYRGNYQGSSTRLAFSKR 154
Db 136 GTRFLHLIFARTLTPEVNPYHGKYQGQGVTLPRK 171

RESULT 6
D72724
probable dCTP deaminase (EC 3.5.4.13) APB0333 [similarity] - Aeropyrum pernix (strain K1
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
A;Accession: D72724
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Rep. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: D72724
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-181 <RAW>
A:Cross-references: UNIPROT:Q9YFA8; DDBJ:AP000059; NID:95103911; PIDN:BA979288.1; PID:95
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0333
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 28.8%; Score 232.5; DB 2; Length 181;
Best Local Similarity 31.8%; Pred. No. 1.2e-14;
Matches 55; Conservative 38; Mismatches 57; Indels 23; Gaps 4;

QY 1 MLLPDKIRK-----EILIEPFSEESLQAGYDLRVGREAFVK-----G 39

Db 4 LILSDRDLALLAIGDLVVEPLSGDITVRENGDLRLGR-GFCRFKRSRDLVDPAPGSPG 62

QY 40 KLIDVEKEGKVIPPREYALILTLERIKLPDDVMGDMKIRSSLAREGVIGSPFAWVDPGWD 99

Db 63 EFVECGEGDEIIVGPEHMLLTQEVIRLPGYAGLVNLRSTWARTGIYIPATVVDAGFE 122

QY 100 GMLTLMYNASNEPVELRYGERFVOIAFIRLEGPARNPYRGNYQGSTRLAFSK 152

Db 123 GQLTIWV-GSGFPVKLYPGDRFLHLVLVKLQSPANNPYRGYQGGVRLPK 174

RESULT 7

T36613
probable dCTP deaminase (EC 3.5.4.13) SCH35.46 [similarity] - Streptomyces coelicolor
C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: T36613

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1999

A:Reference number: Z21610

A:Accession: T36613

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <OLI>

A:Cross-references: UNIPROT:Q9X8W0; EMBL:AL078610; PIDN:CAB44381.1; GSPDB:GN000070; SCOE

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCH35.46

C:Superfamily: dCTP deaminase

C:Keywords: hydrolase

Query Match 26.8%; Score 216; DB 2; Length 191;

Best Local Similarity 31.4%; Pred. No. 4.7e-13;

Matches 53; Conservative 32; Mismatches 60; Indels 24; Gaps 3;

QY 1 MLLPDKIRKEI-----LIEPFSEESLQAGYDLRVGR-----EAFVK 38

Db 1 MLLSDKIDRAEIDNGRVRIDPDDSMVQPSIDVRLDRYFRVFNHRYPHIDPSVEQVLD 60

QY 39 GKLIDVEKEGKVIPPREYALILTLERIKLPDDVMGDMKIRSSLAREGVI--GSPFAWVDP 96

Db 61 TRLVEPEGDEPILHGEFVLASTYEVWLPDDLASRLGKSLGLGLVTHSTAGFIDP 120

QY 97 GWDGNLTLMYNASNEPVELRYGERFVOIAFIRLEGPARNPYRGNYQGS 145

Db 121 GFGSHVTLSNLATLPKLVPMQMKIGQLCFLRLTSPAEHPYGSERYGS 169

RESULT 8

A90160

deoxycytidine triphosphate deaminase (dcD-1) [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 17-May-2002

QY 148 LAFSKRK 154

QY 148 LAFSKRK 154

QY 148 LAFSKRK 154

C:Accession: A90160

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A90139

A:Accession: A90160

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-206 <KUR>

A:Cross-references: GB:AE006641; NID:gl3813326; PIDN:AAK40536.1; GSPDB:GN00155

C:Genetics:

A:Gene: dcd-1

C:Superfamily: dCTP deaminase

Query Match 25.7%; Score 207.5; DB 2; Length 206;

Best Local Similarity 33.5%; Pred. No. 3.2e-12;

Matches 57; Conservative 30; Mismatches 64; Indels 19; Gaps 5;

QY 1 MLLPDKWI-----RKEILIEPFSEESLQAGYDLRVGRE--AFVK-GKLD----- 43

Db 23 MILSDRDLKYLEKSWIKIQLREDTIRENGVLDVRVGNBIARPKTKDIPDNPPSPFF 82

QY 44 -VEKEGKVIPPREYALILTLERIKLPDDVMGDMKIRSSLAREGVIGSPFAWVDPGWDGNL 102

Db 83 QTEKGEEFTIQPYEHVLLTTEYIELNDVMAFVNLRTFARLGLFIPPTIVDAGFGQV 142

QY 103 TLMYNASNEPVELRYGERFVOIAFIRLEGPARNPYRGNYQGSTRLAFSK 152

Db 143 TIEVV-GSFPVKLRKSTRFIHLIFARTLTTPVEYIQQYQGGVTLPK 191

RESULT 9

A69114

dCTP deaminase (EC 3.5.4.13) MTH1847 [similarity] - Methanobacterium thermoautotrophicum

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 31-Mar-2000

C:Accession: A69114

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;

Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func

A:Reference number: A69000; MUID:98037514; PMID:9371463

A:Accession: A69114

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-200 <MTH>

A:Cross-references: GB:AE000937; GB:AE000666; NID:g2622974; PIDN:AAB86313.1; PID:g26229

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1847

C:Superfamily: dCTP deaminase

C:Keywords: hydrolase

Query Match 24.7%; Score 199; DB 2; Length 200;

Best Local Similarity 31.6%; Pred. No. 2e-11;

Matches 59; Conservative 29; Mismatches 63; Indels 36; Gaps 7;

QY 2 LPLDWMKIRKE-----ILIEPFS--BESLQAGYDLRVGREAFVKGLI----- 42

Db 6 ILSDRDLKRYIEGLTITDPLDPERIQIPSSVDLRIGNE--FKGPRVKRKCIDPKDPS 63

QY 43 DVE-----KEGKVIPPREYALILTLERIKLPDDVMGDMKIRSSLAREGVI--GSPA 92

Db 64 DIESYMETPHVEDGPFIIHPGFALATTHEYALPEDLVARVEGRSSIGRLGITHMVTAG 123

QY 93 WYDPGWDGNLTLMYNASNEPVELRYGERFVOIAFIRLEGPARNPYRGNYQGSSTR 147

Db 124 YIDPFGHRTITLISNIGKMPVALYPRQVVCQIVFETMTSPAEHPYGHPSRDSKYIGQTR 183

QY 148 LAFSKRK 154

QY 148 LAFSKRK 154

QY 148 LAFSKRK 154

```
Db          184  POTSRIK 190
|:|
RESULT 10
B70526
dCTP deaminase (EC 3.5.4.13) RV0321 [similarity] - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: B70526
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70526
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-190 <COL>
A:Cross-references: UNIPROT:O07247; GB:Z96800; GB:AL123456; NID:g3261800; PIDN:CAB09605.
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: dcd
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match      24.6%; Score 198; DB 2; Length 190;
Best Local Similarity 30.2%; Pred. No. 2.3e-11;
Matches 51; Conservative 33; Mismatches 61; Indels 24; Gaps 3;

Qy          1  MLLPDWKIRKEIL-----IEPFSESLQPAGYDLRV-----GREAFVK 38
|||
Db          1  MLLSDRLRAEITAGRSIDPFDTLVQPSIDVRLDCMFRVFNTRYTHIDPARQDEL 60
|:|
Qy          39  GKLIDVEKEGKVIPPREYALILTLERIKLPDDVGMKIRSSLAREGVI--GSFAWVDP 96
|:|
Db          61  TSLVQVDPGEFVLHPGFGVLGSLTLEFLTPDLNLAGRLEKSSLGRLGLLTHSTAGFDIP 120
|:|
Qy          97  GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYQGS 145
|:|
Db          121  GFSGHITLSENVANLPITLWPGMKIGQLCMLRLTSPSEHPYGSRRAGS 169
|:|
Qy          97  GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYQGS 145
|:|
Db          121  GFSGHITLSENVANLPITLWPGMKIGQLCMLRLTSPSEHPYGSRRAGS 169
|:|

RESULT 11
H87222
probable deoxycytidine triphosphate deaminase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: H87222
R:Coile, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, S.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, S.
A>Title: Massive gene decay in the leprosy bacillus
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H87222
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-190 <STO>
A:Cross-references: UNIPROT:Q9CB17; GB:AL450380; NID:g13094049; PIDN:CAC32024.1; GSPDB:CAC32024.1
C:Genetics:
A:Gene: ML2507
C:Superfamily: dCTP deaminase

Query Match      24.5%; Score 197.5; DB 2; Length 190;
Best Local Similarity 30.6%; Pred. No. 2.6e-11;
Matches 53; Conservative 32; Mismatches 59; Indels 29; Gaps 4;

Qy          1  MLLPDWKIRKEIL-----IEPFSESLQPAGYDLRV-----GREAFVK 38
|||
|:|
```

```
Db          1  MLLSDRLRAEITAGRSIDPFDTLVQPSIDVRLDCMFRVFNTRYTHIDPARQDEL 60
|:|
Qy          39  GKLIDVEKEGKVIPPREYALILTLERIKLPDDVGMKIRSSLAREGVI--GSFAWVDP 96
|:|
Db          61  TSLVQVDPGEFVLHPGFGVLGSLTLEFLTPDLNLAGRLEKSSLGRLGLLTHSTAGFDIP 120
|:|
Qy          97  GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY-----RGNVQOG 144
|:|
Db          121  GFSGHITLSENVANLPITLWPGMKIGQLCMLRLTSPSEHPYGSRRAGS 173
|:|

RESULT 12
A81272
probable dCTP deaminase (EC 3.5.4.13) Cj1292 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: A81272
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Baaham, D.; Chillingworth, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell, B.G.
Nature 403, 665-668, 2000
A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervirulence
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: A81272
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-186 <PAR>
A:Cross-references: UNIPROT:Q9PN07; GB:AL139078; GB:AL111168; NID:g6968723; PIDN:CAB7371
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: dcd; Cj1292
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match      23.5%; Score 189.5; DB 2; Length 186;
Best Local Similarity 29.4%; Pred. No. 1.4e-10;
Matches 53; Conservative 34; Mismatches 58; Indels 35; Gaps 7;

Qy          5  DWKIRKEIL-----IEPFSEES-----LQAGYDLRVGRE-----AF 36
|:|
Db          7  NW-IRKVALEHKMETEPCANIGKGVSVSYGLSSYGIIVRGREFPKIFTNVNSTVVDPKNF 65
|:|
Qy          37  VKGLIDVEKEGKV-VIPREYALILTLERIKLPDDVGMKIRSSLAREGVI--GSFAWVDP 95
|:|
Db          66  VEENVVDF--EGDVCIVPANSFALARTIEYFPMNDVLAICLGKSTYARCCGIIIVNVTPE 123
|:|
Qy          96  PGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY---RGNVQOGSTRLAFSK 152
|:|
Db          124  PGFEGHITIEISNTTPLPAKIYANEGIAQVLFLOQDEKCDTTYKDKKGKQAQTGITLPR 183
|:|

RESULT 13
G84184
deoxycytidine triphosphate deaminase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: G84184
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Laeky, S.; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jable, J.; Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li, S.; Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Laeky, S.; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jable, J.; Jung, K.H.; Alam, M.; Freitas, T.
A>Title: Genome sequence of Halobacterium species NRC-1
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: G84184
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-195 <STO>
A:Cross-references: UNIPROT:Q9HSG3; GB:AE004437; NID:g10579885; PIDN:AAG18843.1; GSPDB:AAG18843.1
C:Genetics:
A:Gene: dtd
C:Superfamily: dCTP deaminase

Query Match      22.9%; Score 184.5; DB 2; Length 195;
Best Local Similarity 29.4%; Pred. No. 1.4e-10;
Matches 53; Conservative 34; Mismatches 58; Indels 35; Gaps 7;
```

Best Local Similarity 31.2%; Pred. No. 4.5e-10;
Matches 54; Conservative 31; Mismatches 59; Indels 29; Gaps 7;
QY 1 MLPLDKWKIRK-----BILIEPSEESL--QPAGYDLRVGRE--AFVKGKL-----41
| | | | | : | | | | | : | | | | : | | | | :
Db 1 MILSDQDILARLADGLAIEPLDVLQVPASVDVRLGRFLFEPRANVPCHPNREDE 60
| | | | | : | | | | | : | | | | : | | | | :
QY 42 ID-----VEKEGVVIPPREYALILTLERIKLPDVMGDMKIRSSLAAREGV--GSFA 92
| | | | | : | | | | | : | | | | : | | | | :
Db 61 VDEYVTTVEEDGDEFILHPGDFVLGTTTKEVERVEPRDLVAQVEGRSLGLAVVHATAG 120
| | | | | : | | | | | : | | | | : | | | | :
QY 93 WYDPGMDGLTMLVNASNEPVELRYGERFVQIAFIRLEGPARNPYRGVNOGS 145
| | | | | : | | | | | : | | | | : | | | | :
Db 121 FIDPGNGRVTLSENLGNKVPVALTPEMRISQLVFTLSPADRPY-GDERGS 172
| | | | | : | | | | | : | | | | : | | | | :
RESULT 14
D64566
dCTP deaminase (EC 3.5.4.13) HP0372 [similarity] - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Mar-2000
C:Accession: D64566
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: D64566
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-190 <TOM>
A:Cross-references: GB:AE000554; GB:AE000511; NID:g2313475; PIDN:AAD07441.1; PID:g231347
C:Genetics:
A:Start codon: GTG
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase
Query Match 22.5%; Score 181; DB 1; Length 190;
Best Local Similarity 28.0%; Pred. No. 9.2e-10;
Matches 47; Conservative 29; Mismatches 64; Indels 28; Gaps 4;
QY 13 LIEPFSEE-----SLQPAGYDLRVGRE-----AFVKGK-----LIDVEKE 47
| | | | | : | | | | | : | | | | : | | | | :
Db 20 MISPFCEKQGVKNVISYGLSSYGYDIRVGSFPMFPDNKNALIDPKNFDPNNAKIDASK 79
| | | | | : | | | | | : | | | | : | | | | :
QY 48 GKVVIPPREYALILTLERIKLPDVMGDMKIRSSLAAREGVISFAWVDPGMDGNLTMLY 107
| | | | | : | | | | | : | | | | : | | | | :
Db 80 GYFILPANAFALAHTTEYFKMPKDTLAI CLGKSTYARCGIIVNVTPPEPEFGYITIEIS 139
| | | | | : | | | | | : | | | | : | | | | :
QY 108 NASNEPVELRYGERFVQIAFIRLEGPARNPYR---GNYOGSTRLAESK 152
| | | | | : | | | | | : | | | | : | | | | :
Db 140 NTTNLPKAYANEGIAQVVFLQGDCEQSYKDRGGYQGVGITLPLK 187
| | | | | : | | | | | : | | | | : | | | | :
RESULT 15
E71860
dCTP deaminase (EC 3.5.4.13) dcd [similarity] - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: Strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: E71860
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: E71860
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <ARN>
A:Cross-references: UNIPROT:Q9ZKD0; GB:AE001529; GB:AE001439; NID:g4155590; PIDN:AAD0658

A:Experimental source: strain J99
C:Genetics:
A:Gene: dcd
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase
Query Match 22.3%; Score 180; DB 2; Length 188;
Best Local Similarity 28.0%; Pred. No. 1.1e-09;
Matches 47; Conservative 29; Mismatches 64; Indels 28; Gaps 4;
QY 13 LIEPFSEE-----SLQPAGYDLRVGRE-----AFVKGK-----LIDVEKE 47
| | | | | : | | | | | : | | | | : | | | | :
Db 18 MINPCEKQGVKNVISYGLSSYGYDIRVGSFPMFPDNKNALIDPKNFDPNNAKIDASK 77
| | | | | : | | | | | : | | | | : | | | | :
QY 48 GKVVIPPREYALILTLERIKLPDVMGDMKIRSSLAAREGVISFAWVDPGMDGNLTMLY 107
| | | | | : | | | | | : | | | | : | | | | :
Db 78 GYFILPANAFALAHTTEYFKMPKDTLAI CLGKSTYARCGIIVNVTPPEPEFGYITIEIS 137
| | | | | : | | | | | : | | | | : | | | | :
QY 108 NASNEPVELRYGERFVQIAFIRLEGPARNPYR---GNYOGSTRLAESK 152
| | | | | : | | | | | : | | | | : | | | | :
Db 138 NTTNLPKAYANEGIAQVVFLQGDCEQSYKDRGGYQGVGITLPLK 185
| | | | | : | | | | | : | | | | : | | | | :
Search completed: July 2, 2005, 12:25:15
Job time : 8.93023 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 2, 2005, 08:22:22 ; Search time 3629 Seconds
(without alignments)

2082.947 Million cell updates/sec

Title: US-08-957-709A-71

Perfect score: 806

Sequence: 1 MLPLPWKIRKELIEPFSEE.....PYRGNVQGSTRLAFSKRKL 156

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh
-O=/cgn2.1/USPTO.spool/US08957709/runat_01072005.154631.26005/app.query.fasta_1.327
-DB=GenEmbl -QFMT=fastap -SUFFIX=rgc -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US08957709 @CGN 1 1 4200 @runat 01072005.154631.26005 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	806	100.0	471	1	AY066005 Pyrococcus
2	806	100.0	471	1	AY443493 Pyrococcus
3	806	100.0	471	6	BD133238 Pyrococcus
4	806	100.0	471	6	AR535070 Sequence

C	5	806	100.0	13064	1	AE010292	AE010292 Pyrococcus
	6	788	95.3	253505	1	AP000007	AP000007 Pyrococcus
	7	703	87.2	65118	6	BD430793_17	Continuation (18 o
	8	703	87.2	265118	1	CNSPAX06	AJ248288 Pyrococcus
	9	703	87.2	265118	6	AX041922	AX041922 Pyrococcus
	10	543	67.4	740	6	BD133228	BD133228 Polymeras
	11	543	67.4	740	6	AR180965	AR180965 Sequence
	12	543	67.4	740	6	AR207520	AR207520 Sequence
	13	543	67.4	740	6	AR225643	AR225643 Sequence
	14	543	67.4	740	6	AR535060	AR535060 Sequence
	15	257.5	31.9	338100	1	TACID2	AL445064 Thermopla
	16	245.5	30.5	75294	7	AJ783769	AJ783769 Sulfolobu
	17	244	30.3	35450	14	SVI344259	AJ344259 Sulfolobu
	18	243	30.1	14286	1	AE000747	AE000747 Aquifex a
	19	242.5	30.1	296950	1	AP001508	AP001508 Bacillus
	20	242	30.0	3382	1	DALIG	X63438 D.ambivalen
	21	242	30.0	3382	6	A78674	A78674 Sequence 5
	22	237	29.4	32308	14	AF022221	AF022221 Archaeal
	23	237	29.4	32308	14	SVI414696	AJ414696 Sulfolobu
	24	236	29.3	10898	1	AE009852	AE009852 Pyrobacul
	25	233	28.9	543	6	BD180240	BD180240 Highly th
	26	233	28.9	543	6	BD180240	AE017307 Thermus t
	27	233	28.9	247910	1	AE017307	AP000981 Sulfolobu
	28	232.5	28.8	167000	1	AP000059	AP000059 Aeropyrum
	29	232	28.8	110000	1	AE017261	Continuation (2 of
	30	227.5	28.2	110000	1	AE017199_2	Continuation (3 of
	31	220.5	27.4	239340	1	AP000993	AP000993 Thermopla
	32	220	27.3	10375	1	AE010327	AE010327 Methanopy
	33	218.5	27.1	302174	1	AE017241	AE017241 Mycobacte
	34	214	26.6	567	6	BD165341	BD165341 Novel pol
	35	214	26.6	567	6	AX123224	AX123224 Sequence
	36	214	26.6	309400	6	AX127153	AX127153 Sequence
	37	214	26.6	325651	1	AP005283	AP005283 Corynebac
	38	214	26.6	349115	1	BX927156	BX927156 Corynebac
	39	214	26.6	349980	6	AX127152	AX127152 Sequence
	40	212	26.3	309050	1	SC0939117	AL939117 Streptomy
	41	210	26.1	299925	1	AP005039	AP005039 Streptomy
	42	209.5	26.0	110000	1	AP006618	Continuation (58 o
	43	207.5	25.7	11922	1	AE006657	AE006657 Sulfolobu
	44	205.5	25.5	302938	1	BX251412	BX251412 Tropherym
	45	205.5	25.5	324227	1	AE016852	AE016852 Tropherym

ALIGNMENTS

RESULT 1	AY066005	AY066005	Pyrococcus furiosus	dutPase gene, complete cds.	471 bp	DNA	linear	BCT 28-JAN-2002
LOCUS	AY066005	Pyrococcus furiosus	dutPase gene, complete cds.					
DEFINITION	AY066005	Pyrococcus furiosus	dutPase gene, complete cds.					
ACCESSION	AY066005.1	GI:18389121						
VERSION	AY066005.1	GI:18389121						
KEYWORDS								
SOURCE								
ORGANISM								
REFERENCE								
AUTHORS								
TITLE								
JOURNAL								
MEDLINE								
PUBMED								
REFERENCE								
AUTHORS								
TITLE								
JOURNAL								
FEATURES								
Source								

```
CDS
1. 471
/codon_start=1
/transl_table=11
/product="dUTPase"
/protein_id="AAL47572.1"
/db_xref="GI:18389122"
/translation="MLLPDWKIRKEILIEPFSEESLOPAGYDLRVGREAFVKGKLDIV
EKEGKVIPPREYALILTLERIKLPDDVGMKIRSSLRSGVIGSFAMVDPGWDGNL
TLMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL"

ORIGIN
US-08-957-709A-71 (1-471) x AY066505 (1-471)
QY 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuLeuGluProPheSerGluGlu 20
DB 1 ATGCTACTTCCAGACTGGAAATCAGAAAAGAAATCTATAGAGCCATTTCTGAAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTGTAAAGGGGAAA 120
QY 41 LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle 60
DB 121 TTAATCGAGCTGGAAAGAGAAAGAAAGTCTGTTATCTCCAAAGGGAATACGCCCTTAATC 180
QY 61 LeuThrLeuGluArgLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
DB 181 CTAACCTCGAGAGGATAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
QY 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGGTTGGTTCGCCAGGATGGATGGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACTAATGCTCTACATGCTCAATGCTCAATGAAACCTGTCGATTAAGATATGGAG 360
QY 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
DB 361 AGATTTGTCAGATCGATTTATAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLeu 156
DB 421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCAGAGAGAAAGAACTC 468

RESULT 2
AY443493 471 bp DNA linear BCT 17-NOV-2003
LOCUS Pyrococcus woesei dUTPase gene, complete cds.
ACCESSION AY443493
VERSION AY443493.1 GI:38261503
KEYWORDS Pyrococcus woesei
SOURCE Pyrococcus woesei
ORGANISM Pyrococcus woesei
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
1 (bases 1 to 471)
REFERENCE
Dabrowski, S. and Kjaer Ahring, B.
TITLE Cloning, expression, and purification of the His6-tagged
hyper-thermostable dUTPase from Pyrococcus woesei in Escherichia
coli: application in PCR
JOURNAL Protein Expr. Purif. 31 (1), 72-78 (2003)
MEDLINE 22844507
PUBMED 12963343
REFERENCE 2 (bases 1 to 471)

AUTHORS Dabrowski, S. and Ahring, B. K.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-2003) Biocentrum-DTU, Technical University of
Denmark, Soltofte Plads 227, Kgs. Lyngby 2800, Denmark
FEATURES
source
1. 471
/organism="Pyrococcus woesei"
/mol_type="genomic DNA"
/db_xref="taxon:2262"
1. 471
/codon_start=1
/transl_table=11
/product="dUTPase"
/protein_id="AAR15897.1"
/db_xref="GI:38261504"
/translation="MLLPDWKIRKEILIEPFSEESLOPAGYDLRVGREAFVKGKLDIV
EKEGKVIPPREYALILTLERIKLPDDVGMKIRSSLRSGVIGSFAMVDPGWDGNL
TLMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL"

CDS
US-08-957-709A-71 (1-156) x AY443493 (1-471)
QY 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuLeuGluProPheSerGluGlu 20
DB 1 ATGCTACTTCCAGACTGGAAATCAGAAAAGAAATCTATAGAGCCATTTCTGAAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTGTAAAGGGGAAA 120
QY 41 LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle 60
DB 121 TTAATCGAGCTGGAAAGAGAAAGAAAGTCTGTTATCTCCAAAGGGAATACGCCCTTAATC 180
QY 61 LeuThrLeuGluArgLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
DB 181 CTAACCTCGAGAGGATAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
QY 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGGTTGGTTCGCCAGGATGGATGGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACTAATGCTCTACATGCTCAATGCTCAATGAAACCTGTCGATTAAGATATGGAG 360
QY 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
DB 361 AGATTTGTCAGATCGATTTATAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLeu 156
DB 421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCAGAGAGAAAGAACTC 468

RESULT 3
BD133238 471 bp DNA linear PAT 18-SEP-2002
LOCUS Pyrococcus enhancing factor (PEF) extracts, PEF protein complexes,
isolated PEF proteins, and methods for purifying and identifying
them.
DEFINITION BD133238
ACCESSION BD133238
VERSION BD133238.1 GI:23228183
KEYWORDS JP 200250572-A/34.
SOURCE unidentified
ORGANISM unidentified
```

unclassified.
1 (bases 1 to 471)
REFERENCE Hogrefe,H. and Hansen,C.J.
AUTHORS Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
TITLE isolated PEF proteins, and methods for purifying and identifying
JOURNAL Patent: JP 2002505572-A 34 19-FEB-2002;
STRATAGENE

COMMENT
OS Unidentified
PN JP 2002505572-A/34
PD 19-FEB-2002
PF 20-MAR-1998 JP 1998545828
PR 21-MAR-1997 US 08/822774,24-OCT-1997 US 08/957709 PI
PC HOLLY HOGREFE, CONNIE J HANSEN
PC C12P19/34, C12Q1/68 C12N9/12
CC Strandedness: Single;
CC Topology: Linear;
CC Polymerase enhancing factor (PEF) extracts, PEF protein CC
complexes,
CC isolated PEF proteins, and methods for purifying and CC
identifying them

PH Key Location/Qualifiers
FT source 1..471
FT /organism='Unidentified'.
FT Location/Qualifiers

FEATURES
source
1..471
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN

Alignment Scores:
Pred. No.: 1.04e-80 Length: 471
Score: 806.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x BD133238 (1-471)

QY 1 MetLeuLeuProAspTrrpLysileArgLysGluLeuLeuGluProPheSerGluGlu 20
DB 1 ATGCTACTTCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTCTTAAGGGGAAA 120
QY 41 LeuLeuAspValGluLysGluGlyLysValValleProPheArgGluTyrAlaLeulle 60
DB 121 TTAATCGACGTGAAAAGGAGGAAAGTTCGTTATTCCTCCAAAGGGAATACGCTTAATC 180
QY 61 LeuThrLeuGluArgLileLysLeuProAspAspValMetGlyAspMetLysileArgSer 80
DB 181 CTAACCTCGAGAGGATAAAGTTGCCCGACGATGTTATGGGGATATGAAGATAGAGGC 240
QY 81 SerLeuAlaArgGluGlyValleGlySerPheAlaTrrpValAspProGlyTrrpAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTTATGCTTTCTTTTGGTTGACCCAGGATGGATGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACACTAATGCTCTACAAATGCCTCAATGAACCTGTGCAATTAAGATATGAGAG 360
QY 121 ArgPheValGlnleAlaPheleArgLeuGluGlyProAlaArgAsnProTrrpArgGly 140
DB 361 AGATTGTGCAGATCGCATTTATAAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLysLeu 156
DB 421 AACTATCAGGGAGGCACAGGTTAGCGTTTTCAGAGAGAAAGAACTC 468

RESULT 4

Archaea; Euryarchaeota; Thermococci; Thermococcales;

AR535070
LOCUS AR535070 471 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 70 from patent US 6734293.
ACCESSION AR535070
VERSION AR535070.1 GI:53925890
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 471)
AUTHORS Hogrefe,H. and Hansen,C.J.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
JOURNAL isolated PEF protein, and methods for purifying and identifying
FEATURES Patent: US 6734293-A 70 11-MAY-2004;
Location/Qualifiers
1..471
source /organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1.04e-80 Length: 471
Score: 806.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x AR535070 (1-471)

QY 1 MetLeuLeuProAspTrrpLysileArgLysGluLeuLeuGluProPheSerGluGlu 20
DB 1 ATGCTACTTCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTCTTAAGGGGAAA 120
QY 41 LeuLeuAspValGluLysGluGlyLysValValleProPheArgGluTyrAlaLeulle 60
DB 121 TTAATCGACGTGAAAAGGAGGAAAGTTCGTTATTCCTCCAAAGGGAATACGCTTAATC 180
QY 61 LeuThrLeuGluArgLileLysLeuProAspAspValMetGlyAspMetLysileArgSer 80
DB 181 CTAACCTCGAGAGGATAAAGTTGCCCGACGATGTTATGGGGATATGAAGATAGAGGC 240
QY 81 SerLeuAlaArgGluGlyValleGlySerPheAlaTrrpValAspProGlyTrrpAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTTATGCTTTCTTTTGGTTGACCCAGGATGGATGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACACTAATGCTCTACAAATGCCTCAATGAACCTGTGCAATTAAGATATGAGAG 360
QY 121 ArgPheValGlnleAlaPheleArgLeuGluGlyProAlaArgAsnProTrrpArgGly 140
DB 361 AGATTGTGCAGATCGCATTTATAAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLysLeu 156
DB 421 AACTATCAGGGAGGCACAGGTTAGCGTTTTCAGAGAGAAAGAACTC 468

RESULT 5

AE010292/c

LOCUS AE010292 13064 bp DNA linear BCT 25-FEB-2002
DEFINITION Pyrococcus furiosus DSM 3638, section 167 of the complete
genome.

ACCESSION AE010292 AB009950

VERSION AE010292.1 GI:18894190

KEYWORDS

SOURCE Pyrococcus furiosus DSM 3638

ORGANISM Pyrococcus furiosus DSM 3638

Archaea; Euryarchaeota; Thermococci; Thermococcales;

RFKLEVLPHGRGKDVKIADVAGVAEAAARKLGLDVISSAELEETASSPQAROKLAN
 KYDFPIAEABLPKIGRLGRYLGNPNMPTTLDLTPIVEKLKKTIRIOLKNN
 PVVHAPVGTGKMSDEETAENIEAVLNIAIIGKLGRGESQVKSVYVKTMTGPAVKIEG
 5520. .6539
 /gene="PF1993"
 5520. .6539
 /gene="PF1993"
 /note="Function Code: 13.4 Translation: Ribosomal
 proteins: synthesis and modification; (rpl10E)"
 /codon_start=1
 /transl_table=11
 /product="LSU ribosomal protein L10E"
 /protein_id="AAU82117.1"
 /db_xref="GI:18894199"
 /translation="MAHVAEWKKKEVEELANLKSYPVVALVDVSMPAYPLSQMRRL
 IREANGLRYSRNTLIREAKTKVAQELKPELSKLINIEGGAGILVTMNPFLKVF
 LQONRQPAKPAKGVKDVVIPAGPTSLAPGPIVQMGQAMGIPARIERGKVIQKDT
 VVLKAGEIPELANILNAGIOPLEVGLDOLLAYVEDGILYTPDVLAIDESEVINMLQ
 KAYHARNFLAVNTAYPTPTQIEALIQKAFLNAKAVAEGYITKETISDIIIGRAIRAM
 LLQAQPEVDLDEKTKELLUSAQAQVSVAQVEEKKKEEKEEKEEAESEALAGL
 SALFG"
 5586. .6909
 /gene="PF1994"
 5586. .6909
 /gene="PF1994"
 /note="Function Code: 13.4 Translation: Ribosomal
 proteins: synthesis and modification; (rpl12A)"
 /codon_start=1
 /transl_table=11
 /product="LSU ribosomal protein L12A"
 /protein_id="AAU82118.1"
 /db_xref="GI:18894200"
 /translations="MEYVYAALLHSVGKEINEENLKAVLQAAAGVPEDEARIKALVAA
 LEGWNIDEVIEKAAMPVAVAAAPAAAPAGGGEKKKEEKEEKEEVEEAEALAGL
 SALFG"
 7014. .7556
 /gene="PF1995"
 7014. .7556
 /gene="PF1995"
 /note="Function Code: 16.1 Conserved Hypothetical"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAU82119.1"
 /db_xref="GI:18894201"

Alignment Scores:
 Pred. No.: 4,86e-79 Length: 13064
 Score: 806.00 Matches: 156
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 1 Gaps: 0

US-08-957-709A-71 (1-156) x AE010292 (1-13064)

Qy	1	MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu	20
Db	8006	ATGCTACTTCAGACTCGAANAATCAGAAAGAATACTTATAGAGCATTTTCTGAGAA	7947
Qy	21	SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluLaPheValLysGlyLys	40
Db	7946	TCGCTCAACACAGCAGGTTATGACCTCAGAGTGGCAGAGAGGCTTTGTTAAGGGGAA	7887
Qy	41	LeuIleAspValGluLysGluLysValValIleProProArgGluTyrAlaLeuIle	60
Db	7886	TTAATCACCTGGAAAAAGGAAGGAAAGTCTGTTATTCCTCAAGGGAATACGCTTAATC	7827
Qy	61	LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer	80
Db	7826	CTAACCTTCAGAGAGATAAAGTTGCCCGACGATGTTATGGGGGATATGAAGATAAGAGC	7767
Qy	81	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly	100

Db	7766	AGTTTACAGAGAAGGGGTATTGGTTCCTTTTGGCTTGGGTGACCCAGGATGGATGGA	7707
Qy	101	AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120
Db	7706	AACCTTAACACTAATGCTCTCAATGCCTCAATGAACCTGTCGAATTAAGATATGAGAG	7647
Qy	121	ArgPheValGlnIleAlaPheLeuArgLeuGluGlyProAlaArgAsnProTyrArgGly	140
Db	7646	AGATTTTGTGCAGATCGCATTTATATAAGGCTAGAGGTCGCGCAAGAAACCTTTACAGAG	7587
Qy	141	AsnTyrGlnGlySerThrArgLeuAlaPheSerIleArgIleValysLeu	156
Db	7586	AACATATCAGGGGAGCACAGGTTAGCGTTTTCAAAGAGAAAGAAATC	7539
RESULT 6			
AP000007		253505 bp	DNA linear BCT 27-MAY-2004
LOCUS			
DEFINITION	Pyrococcus horikoshii OT3 DNA, complete genome, 1485001-1738905 nt		
ACCESSION	AP000007 AB009464 AB009465 AB009521 AB009522 AB009523 AB009524		
	AB009525 AB009526 AB009527 AB009528 AB009529 AB009530 AB009531		
VERSION	BA000001		
KEYWORDS	AP000007.1 GI:3236134		
SOURCE	Pyrococcus horikoshii OT3		
ORGANISM	Pyrococcus horikoshii OT3		
REFERENCE	Archaea: Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus.		
AUTHORS	1		
	Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamida, N., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Nakamura, Y., Robb, T. F., Horikoshi, K., Masuchi, Y., Shizuya, H. and Kikuchi, H.		
TITLE	Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3		
JOURNAL	DNA Res. 5 (2), 55-76 (1998)		
MEDLINE	98344137		
PUBMED	9679194		
AUTHORS	2. (bases 1 to 253505)		
	Director-General of Biotechnology Center, Tanaka, T., Kawarabayasi, Y. and Kikuchi, H.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-JUN-1998) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan (E-mail: bio@nitech.go.jp, Tel:81-3-3481-1933, Fax:81-3-3481-8424) On or before Mar 17, 1999 this sequence version replaced gi:3131896, gi:3131920, gi:3131974, gi:3131982, gi:3132032, gi:3132063, gi:3132118, gi:3132139, gi:3132184, gi:3132233, gi:3133342.		
COMMENT	Kawarabayasi, Y. is officially affiliated with the National Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki 305-0046, Japan. Robb, T. F. is at the Center of Marine Biotechnology, University of Maryland, Baltimore, MD, USA. Horikoshi, K. is at the Japan Marine Science and Technology Center, Yokosuka, Kanagawa 237-0061, Japan. Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041, Japan. Shizuya, H. is at the California Institute of Technology, Pasadena, CA, USA. The other authors are at the National Institute of Technology and Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan. All the sequence with length 100 codons or more between ATG or GTG and stop codon are defined as CDS Homology analysis is performed by Smith-Waterman algorithm against GenBank and GenPept release 103; EMBL release 52.0; SwissProt release 34.0; PIR-Protein release 54.0; and OML release 29.5. E-mail address for comments and questions: genome@nitech.go.jp Restriction map, ORF organization, sequence alignment and more		

information are available at w.w.w. site of Biotechnology Center,
URL: <http://www.bio.nits.go.jp/>.

FEATURES

source
1..253505
/organism="Pyrococcus horikoshii OT3"
/mol_type="genomic DNA"
/strain="OT3"
/db_xref="taxon:70601"
/map="1485001-1738505 nt. position"

ORIGIN

Alignment Scores:
Pred. No.: 2 648-73 Length: 253505
Score: 768.00 Matches: 146
Percent Similarity: 96.15% Conservative: 4
Best Local Similarity: 93.59% Mismatches: 6
Query Match: 95.29% Indels: 0
DB: 1 0 Gaps: 0

US-08-957-709A-71 (1-156) x AP000007 (1-253505)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluLeuLeuGluProPheSerGluGlu 20
Db 250855 ATGCTACTTCGGAGCTGGAAATAAGAAAGAAATACCTTATAGAACCACTTTCTCGAAGAA 250914
Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
Db 250915 TCACCTTCAACACGACGAGGTATGATCTTAGAGTGGGTAGAGAAGCTTTTGTAGTCGAAAG 250974
Qy 41 LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle 60
Db 250975 TTAATGTGATGGAAAGAGGGAAGTAGTATTATCTCCAGGGAATACGCTCTAATT 251034
Qy 61 LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleArgSer 80
Db 251035 CTAACCTCGAGAGAGTAAAGTTACTGTATGATGTGATGGGAGATATGAAGATAAGGAGC 251094
Qy 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100
Db 251095 AGCTTAGCTAGAGAGGGATTCTAGGTTCTTTTGGCTGGTTGACCCAGGATGGGATGGG 251154
Qy 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
Db 251155 AACTTAACCTTAATGCTCTACAAATGCTTCAATGACCCAGTACGATTAATTAATACCGAGNA 251214
Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
Db 251215 AGGTTTGTGCAGATTGCTTTTATAGGCTAGAGGATCCACCAAGGAATCCCTATAGTGA 251274
Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLeu 156
Db 251275 AACTATCAGGAAGCACAGATTGGTATTATTTCCAAAGAGAAGAAACTC 251322

RESULT 7

BD430793_17

WPCOMMENT

Sequence split into 18 fragments LOCUS BD430793 Accession BD430793

Fragment Name	Begin	End
BD430793_00	1	110000
BD430793_01	100001	210000
BD430793_02	200001	310000
BD430793_03	300001	410000
BD430793_04	400001	510000
BD430793_05	500001	610000
BD430793_06	600001	710000
BD430793_07	700001	810000
BD430793_08	800001	910000
BD430793_09	900001	1010000
BD430793_10	1000001	1110000
BD430793_11	1100001	1210000
BD430793_12	1200001	1310000
BD430793_13	1300001	1410000
BD430793_14	1400001	1510000

BD430793_15 1500001 1610000
BD430793_16 1600001 1710000
BD430793_17 1700001 1765118

Continuation (18 of 18) of BD430793 from base 1700001 (BD430793 GENOME SEQUENCE AND POLY

Alignment Scores:

Pred. No.: 9 91e-67 Length: 65118
Score: 703.00 Matches: 131
Percent Similarity: 94.12% Conservative: 13
Best Local Similarity: 85.62% Mismatches: 9
Query Match: 87.22% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x BD430793_17 (1-65118)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluLeuLeuGluProPheSerGluGlu 20
Db 62045 ATGCTCTTCAGACTCGAAATAAGGAAGAGATTTTAAATAAAGCCATTTCTCAGAGGAA 62104
Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
Db 62105 TCGCTTCAACACGAGCTGGCTAGACTTAAGGGTAGCAAGGAAGCTTACATTCGGGAAG 62164
Qy 41 LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle 60
Db 62165 TTCATAGATGTTGAAAGAGGAGGCAAGCTCATATACCTCCAAAGAATATGCCCTAATA 62224
Qy 61 LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleArgSer 80
Db 62225 CTGACCTTAGAGAGGATAAAGCTTCCAGATGATATTATGGGAGACATGAAGATAAGGAGC 62284
Qy 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100
Db 62285 AGCTAGCTAGGAGGAGGTGTTCTAGGCTCCTTTGTCATGGGTAGACCCGGGATGGACGGT 62344
Qy 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
Db 62345 AATCTAACCTTGATGCTTTACAAATGCAATCGAAAGGAGGTAATTTTAAGGTACAAAGAG 62404
Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
Db 62405 AGGTTTGTCCAGATAGCCTTCTTAAGGCTTGAGGCTCCCGCAAGAATCCATACAGGGC 62464
Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArg 153
Db 62465 AACTATCAAGGAAGCAGGAGAATAGTCTCTCAAAAGAGA 62503

RESULT 8

CNSPAX06

LOCUS

Pyrococcus abyssi

complete genome; segment 6/6.

Accession

AJ248288.1

GI:5458960

complete genome.

Pyrococcus abyssi

Pyrococcus abyssi

Archaea; Euryarchaeota; Thermococci; Thermococcales;

Thermococcaceae; Pyrococcus.

1 (bases 1 to 265118)

Gaspin,C., Cavaillie,J., Erauso,G. and Bachelierie,J.P.

Archaeal homologs of eukaryotic methylation guide small nucleolar

RNA: lessons from the Pyrococcus genomes

J. Mol. Biol. 297 (4), 895-906 (2000)

20202371

10736225

REFERENCE

AUTHORS

Lecompte,O., Ripp,R., Puzos-Barbe,V., Duprat,S., Heilig,R.,

Districh,J., Thierry,J.C. and Poch,O.

Genome evolution at the genus level: comparison of three complete

Genomes of hyperthermophilic archaea

Genome Res. 11 (6), 981-993 (2001)

21275479

PUBMED

11381026


```
APLEKLKYAQAALAEGRLEKQVSEYLRQIRYLRLDERDEIGALIOAFBAVSXDLVGTNLNALS
KKLERLAEGLDLSNGLTVRGELRDILIEDIRSVTETPRESIGSLVEMANDLEKRANL
AQVSKDVTAINQVNEALQQVSIERQROQETINEITDGMRLVAQTSSEVRAMFEFSG
AVTEVDVTSNGSQDALKRDIEDIQHMSRIETBKVAMERNSIBEITNVITSIA
EQTNLLANAAIEAARAGEAGRFVAQEIIRKLAEBSKQADNIKSIIDITDIKE
AVEATKGVSVIGESSETRDTIGVLANIATILLOETSERMTTVKEQIVRTQREVDKAL
RALENLAASABETTAAAEVSSAIEBQTTAAIEELRRAQELKDMVGRMRQIVGRPKV"
complement(9704..10195)
CDS
/note="PAB1335"
/codon_start=1
/transl_table=11
/product="cheD chemotaxis protein cheD"
/protein_id="CAB50455.1"
/db_xref="GI:5458969"
/db_xref="UniProt/TREMBL:Q9UJF7"
/translation="MTRIKVIGIDYAVGKGEGIIYSTVLGSCVGIITLYDVRTVKVGG
LHALLPEAARGHGRNPARYVDITGLQLLEVLKLGASKFRLEAKLFGQAMQFONIKS
DELKIGERNVQAKRELKGLIRLVAEDTGGRGRTIYLDLSTGKVRMRKVIIGGVIE
KVY"
complement(10179..10817)
CDS
/note="PAB1334"
/codon_start=1
/transl_table=11
/product="cheC-1 chemotaxis protein cheC"
/protein_id="CAB50456.1"
/db_xref="GI:5458970"
/db_xref="UniProt/TREMBL:Q9UJF6"
/translation="MKKSELYKIDFKESANIAMSHALTSLSQMGIGPIEMSAAPDVEIV
SRAEFLKLAERGSGFTVMFDVTEGLSGITLIQFPKXSAINISAVLIMGHEPSMEE
LDEMKSAIMVEGNILISAYTDILSNLIGEPVSLSPKPAESLVDIEKELGRPDLRNV
Alignment Scores:
Pred. No.: 5,04e-66 Length: 265118
Score: 703.00 Matches: 131
Percent Similarity: 94.12% Conservative: 13
Best Local Similarity: 85.62% Mismatches: 9
Query Match: 87.22% Indels: 0
DB: 1 Gaps: 0
US-08-957-709A-71 (1-156) x CNSPAX06 (1-265118)
QY 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuLeuGluProPheSerGluGlu 20
DB 262045 ATGCTCTTCAGACTGGAAATAGGAAGAGATTTTAATAAGCCATTCTCAGAGAA 262104
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 262105 TCGCTTCAACCACTGGCTACGACTTAAGGTTAGGCAAGGAGCTTACATTCAAGGAAAG 262164
QY 41 LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAlaLeuIle 60
DB 262165 TTCATAGATGTTGAAAGAGAGGCAAAAGTCATAATACCTCCAAAAGAAATATGCCCTAATA 262224
QY 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
DB 262225 CTGACCTTAGAGAGGATAAAGCTTCAGATGATATATATGGGAGACATGAGATAGGAGC 262284
QY 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTyrValAspProGlyTrpAspGly 100
DB 262285 AGCTAGCTAGGAAAGGTCTTCTAGGCTCTTTTCATGGTGTAGACCCGGGATGGACGGT 262344
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 262345 AATCTAACCTTGATGCTTTTACAAATGCATCGGAAAGAGGAGTAAATTTTAAGGTACAAAGAG 262404
QY 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
DB 262405 AGGTTGTCCAGATAGCCCTTCTTAGGCTTTAGGCTCCCGCCCAAGAAATCCATACAGGGGC 262464
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArg 153
DB 262465 AACTATCAAGGAAGCAGGAGGAATAGTCTCTCAAGAGA 262503
```

```
RESULT 9
AX041922 265118 bp DNA linear PAT 23-NOV-2000
LOCUS Sequence 817 from Patent WO0065062.
DEFINITION AX041922
ACCESSION AX041922
VERSION AX041922.1 GI:11340687
KEYWORDS Pyrococcus abyssi
SOURCE Pyrococcus abyssi
ORGANISM Pyrococcus abyssi
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
REFERENCE
1 Forterre,P., Thierry,J.C., Prieur,D., Dietrich,J., Lecompte,O.,
Querrelou,J., Weissenbach,J., Saurin,W., Heilig,R., Flament,D.,
Raffin,J.P., Henneke,G., Gueguen,Y. and Rolland,J.L.
Genome sequence and polypeptides of pyrococcus _i(abyssi), fragment
and uses thereof
PATENT: WO 0065062-A 817 02-NOV-2000;
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IFREMER
INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)
FEATURES
Location/Qualifiers
source
1..265118
/organism="Pyrococcus abyssi"
/mol_type="unassigned DNA"
/db_xref="taxon:29292"
/note="Original length of seq 1: 1765118 (for 812
sequences) New sequences : seq 001, from 0.000.001 to
0.349.980, length: 349.980 seq 813, from 0.300.001 to
0.649.980, length: 349.980 seq 814, from 0.600.001 to
0.949.980, length: 349.980 seq 815, from 0.900.001 to
1.249.980, length: 349.980 seq 816, from 1.200.001 to
1.549.980, length: 349.980 seq 817, from 1.500.001 to
1.765.118, length: 265.118"
ORIGIN
Alignment Scores:
Pred. No.: 5,04e-66 Length: 265118
Score: 703.00 Matches: 131
Percent Similarity: 94.12% Conservative: 13
Best Local Similarity: 85.62% Mismatches: 9
Query Match: 87.22% Indels: 0
DB: 1 Gaps: 0
US-08-957-709A-71 (1-156) x AX041922 (1-265118)
QY 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuLeuGluProPheSerGluGlu 20
DB 262045 ATGCTCTTCAGACTGGAAATAGGAAGAGATTTTAATAAGCCATTCTCAGAGAA 262104
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 262105 TCGCTTCAACCACTGGCTACGACTTAAGGTTAGGCAAGGAGCTTACATTCAAGGAAAG 262164
QY 41 LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAlaLeuIle 60
DB 262165 TTCATAGATGTTGAAAGAGAGGCAAAAGTCATAATACCTCCAAAAGAAATATGCCCTAATA 262224
QY 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
DB 262225 CTGACCTTAGAGAGGATAAAGCTTCAGATGATATATATGGGAGACATGAGATAGGAGC 262284
QY 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTyrValAspProGlyTrpAspGly 100
DB 262285 AGCTAGCTAGGAAAGGTCTTCTAGGCTCTTTTCATGGTGTAGACCCGGGATGGACGGT 262344
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 262345 AATCTAACCTTGATGCTTTTACAAATGCATCGGAAAGAGGAGTAAATTTTAAGGTACAAAGAG 262404
QY 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
DB 262405 AGGTTGTCCAGATAGCCCTTCTTAGGCTTTAGGCTCCCGCCCAAGAAATCCATACAGGGGC 262464
```



```

QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArg 153
Db 262465 AACTATCAAGGAAGCAGGAGATAGTCTCTCAAGAGA 262503

RESULT 10
BD133228
LOCUS
DEFINITION
Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
isolated PEF proteins, and methods for purifying and identifying
them.
ACCESSION
BD133228
VERSION
BD133228.1 GI:23228173
KEYWORDS
JP 2002505572-A/24.
SOURCE
unidentified
ORGANISM
unclassified.
REFERENCE
1 (bases 1 to 740)
AUTHORS
Hogrefe, H. and Hansen, C.J.
TITLE
Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
isolated PEF proteins, and methods for purifying and identifying
Patent: JP 2002505572-A 24 19-FEB-2002;
JOURNAL
STRATAGENE
COMMENT
OS Unidentified
PN JP 2002505572-A/24
PD 19-FEB-2002
PF 20-MAR-1998 JP 1998545828
PR 21-MAR-1997 US 08/822774,24-OCT-1997 US 08/957709 P1
HOLLY HOGREFE, CONNIE J HANSEN
PC C12P19/34,C12Q1/68,C12N9/12
CC Strandedness: Single;
CC Topology: Linear;
CC Polymase enhancing factor (PEF) extracts, PEF protein CC
complexes,
CC isolated PEF proteins, and methods for purifying and CC
identifying them
FH Key Location/Qualifiers
FT source 1..740
FT /organism='Unidentified'.
FEATURES
source
Location/Qualifiers
1..740
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'

ORIGIN
Alignment Scores:
Pred. No.: 4,07e-51 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x BD133228 (1-740)

QY 53 ProProArgGluTyrAlaLeuLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 2 CTTCAAGGGATACGCCTTAATCTTAACCTCGAGAGGATTAAGTTGCCGACGATGT 61
QY 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 62 ATGGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAGGGGTATTGTTCTTTTGTCT 121
QY 93 TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
Db 122 TGGGTGACCCAGGATGGATGGAACCTTAACACTAATGCTCTACAAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
Db 182 CTTGTCGAATTGAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAGGCTAGAGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 242 CCGGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAGGTTAGCGTTTCAAG 301

QY 153 ArgLysLysLeu 156
Db 302 AGAAGAAACTC 313

RESULT 12
AR207520
LOCUS
DEFINITION
Sequence 42 from patent US 6379553.
ACCESSION
AR207520
VERSION
AR207520.1 GI:21507294
KEYWORDS
SOURCE
Unknown.
ORGANISM
Unknown.
REFERENCE
1 (bases 1 to 740)
AUTHORS
Hogrefe, H.
TITLE
Polymerase enhancing factor (PEF) extracts, PEF protein complexes,

```

```

Db 242 CCGGCAAGAAACCCCTTACAGAGAAACTATCAGGGGAGCACAAGGTTAGCGTTTCAAG 301
QY 153 ArgLysLysLeu 156
Db 302 AGAAGAAACTC 313

RESULT 11
AR180965
LOCUS
DEFINITION
Sequence 42 from patent US 6333165.
ACCESSION
AR180965
VERSION
AR180965.1 GI:20222298
KEYWORDS
Unknown.
SOURCE
Unknown.
ORGANISM
Unclassified.
REFERENCE
1 (bases 1 to 740)
AUTHORS
Hogrefe, H.
TITLE
Methods for identifying polymerase enhancing factor (PEF)
JOURNAL
Patent: US 6333165-A 42 25-DEC-2001;
FEATURES
Location/Qualifiers
1..740
/organism='unknown'
/mol_type='unassigned DNA'

ORIGIN
Alignment Scores:
Pred. No.: 4,07e-51 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x AR180965 (1-740)

QY 53 ProProArgGluTyrAlaLeuLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 2 CTTCAAGGGATACGCCTTAATCTTAACCTCGAGAGGATTAAGTTGCCGACGATGT 61
QY 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 62 ATGGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAGGGGTATTGTTCTTTTGTCT 121
QY 93 TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
Db 122 TGGGTGACCCAGGATGGATGGAACCTTAACACTAATGCTCTACAAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
Db 182 CTTGTCGAATTGAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAGGCTAGAGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 242 CCGGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAGGTTAGCGTTTCAAG 301

QY 153 ArgLysLysLeu 156
Db 302 AGAAGAAACTC 313

RESULT 12
AR207520
LOCUS
DEFINITION
Sequence 42 from patent US 6379553.
ACCESSION
AR207520
VERSION
AR207520.1 GI:21507294
KEYWORDS
SOURCE
Unknown.
ORGANISM
Unknown.
REFERENCE
1 (bases 1 to 740)
AUTHORS
Hogrefe, H.
TITLE
Polymerase enhancing factor (PEF) extracts, PEF protein complexes,

```

isolated PEF proteins, and methods for purifying and identifying same

JOURNAL Patent: US 6379553-A 42 30-APR-2002;

FEATURES Location/Qualifiers

source

1..740

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:

Pred. No.: 4.07e-51 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x AR207520 (1-740)

Qy 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 2 CCTCCAAGGGAATACGCCTTAATCTAACCCTCGAGAGGATAAAGTTGCCCGACGATGTT 61
Qy 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 62 ATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGGGGTTATTGGTCTTTTGTCT 121
Qy 93 TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
Db 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACAAATGCCCTCAAAATGAA 181
Qy 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
Db 182 CCTGTCGAATTAAGATATGGAGAGAGATTTGTGCAGATCGCATTTATTAAGGCTAGAGGGT 241
Qy 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 242 CCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAAGTTTAGCGTTTTCAAAG 301
Qy 153 ArgLysLysLeu 156
Db 302 AGAAGAAACTC 313

RESULT 13

AR225643 AR225643 740 bp DNA linear PAT 20-DEC-2002

LOCUS Sequence 42 from patent US 644428.

DEFINITION AR225643

ACCESSION AR225643

VERSION AR225643.1 GI:27263675

KEYWORDS

Source

Unknown.

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 740)

AUTHORS

Hogrefe, H.

Polymerase enhancing factor (PEF) extracts, PEF protein complexes,

isolated PEF proteins, and methods for purifying and identifying

same

JOURNAL Patent: US 644428-A 42 03-SEP-2002;

FEATURES Location/Qualifiers

source

1..740

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 4.07e-51 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x AR225643 (1-740)

Qy 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 2 CCTCCAAGGGAATACGCCTTAATCTAACCCTCGAGAGGATAAAGTTGCCCGACGATGTT 61
Qy 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 62 ATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGGGGTTATTGGTCTTTTGTCT 121
Qy 93 TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
Db 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACAAATGCCCTCAAAATGAA 181
Qy 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
Db 182 CCTGTCGAATTAAGATATGGAGAGAGATTTGTGCAGATCGCATTTATTAAGGCTAGAGGGT 241
Qy 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 242 CCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAAGTTTAGCGTTTTCAAAG 301
Qy 153 ArgLysLysLeu 156
Db 302 AGAAGAAACTC 313

RESULT 14

AR535060 AR535060 740 bp DNA linear PAT 08-OCT-2004

LOCUS Sequence 42 from patent US 6734293.

DEFINITION AR535060

ACCESSION AR535060

VERSION AR535060.1 GI:53925880

KEYWORDS

Source

Unknown.

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 740)

AUTHORS

Hogrefe, H. and Hansen, C.J.

Polymerase enhancing factor (PEF) extracts, PEF protein complexes,

isolated PEF protein, and methods for purifying and identifying

Patent: US 6734293-A 42 11-MAY-2004;

JOURNAL Location/Qualifiers

FEATURES source

1..740

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 4.07e-51 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x AR535060 (1-740)

Qy 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 2 CCTCCAAGGGAATACGCCTTAATCTAACCCTCGAGAGGATAAAGTTGCCCGACGATGTT 61
Qy 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 62 ATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGGGGTTATTGGTCTTTTGTCT 121
Qy 93 TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
Db 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACAAATGCCCTCAAAATGAA 181
Qy 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
Db 182 CCTGTCGAATTAAGATATGGAGAGAGATTTGTGCAGATCGCATTTATTAAGGCTAGAGGGT 241

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:26:04 ; Search time 30.9767 Seconds
(without alignments)
1942.879 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKELIIPFSEE.....PYRGNVQSTRLAFSKKKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1722976 seqs, 385795295 residues

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTU5_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	218	27.0	189	9	US-09-738-626-6640
2	210	26.1	191	14	Sequence 6640, Ap
3	152	18.9	195	16	US-10-156-761-11998
4	138	17.1	190	15	Sequence 11998, A
5	134.5	16.7	177	10	US-10-437-963-116832
6	121	15.0	147	17	Sequence 412, App
7	121	15.0	158	18	US-09-975-719-247
8	112	13.9	181	10	Sequence 247, App
9	111	13.8	164	15	US-10-472-928-4752
10	111	13.8	164	15	Sequence 3527, Ap
11	111	13.8	229	15	US-09-975-719-239
					Sequence 239, App
					Sequence 32, Appl
					Sequence 409, App
					Sequence 10, Appl

12	111	13.8	229	15	US-10-336-472-30	Sequence 30, Appl
13	111	13.8	252	15	US-10-136-728-8	Sequence 8, Appl
14	111	13.8	252	15	US-10-336-472-28	Sequence 28, Appl
15	111	13.8	252	16	US-10-408-765A-2116	Sequence 2116, Ap
16	102	12.7	178	9	US-09-970-711-2	Sequence 2, Appl
17	99.5	12.3	254	10	US-09-975-719-241	Sequence 241, App
18	97	12.0	148	16	US-10-627-592-10	Sequence 10, Appl
19	97	12.0	150	16	US-10-627-592-8	Sequence 8, Appl
20	93	11.5	171	16	US-10-437-963-173883	Sequence 173883, Ap
21	90.5	11.2	171	16	US-10-425-115-256676	Sequence 256676, Ap
22	90	11.2	222	16	US-10-739-930-9568	Sequence 9568, Ap
23	89.5	11.1	175	15	US-10-424-599-269427	Sequence 269427, Ap
24	87.5	10.9	1150	9	US-09-946-239-9	Sequence 9, Appl
25	87.5	10.9	1259	15	US-10-282-122A-56089	Sequence 56089, A
26	84	10.4	701	15	US-10-282-122A-55632	Sequence 55632, A
27	82	10.2	863	15	US-10-369-493-3655	Sequence 3655, Ap
28	82	10.2	884	17	US-10-732-923-23250	Sequence 23250, A
29	81.5	10.1	902	14	US-10-231-778-222	Sequence 222, App
30	81	10.0	154	9	US-09-712-363-250	Sequence 250, App
31	80	9.9	149	9	US-09-738-626-5594	Sequence 5594, Ap
32	79.5	9.9	502	14	US-10-006-852-18	Sequence 18, Appl
33	79.5	9.9	502	17	US-10-732-923-6235	Sequence 6235, Ap
34	79	9.8	267	15	US-10-369-493-20803	Sequence 20803, A
35	78.5	9.7	457	15	US-10-369-493-11845	Sequence 11845, A
36	78	9.7	201	10	US-09-988-067B-50	Sequence 50, Appl
37	78	9.7	290	16	US-10-437-963-120944	Sequence 120944, Ap
38	76.5	9.5	146	15	US-10-335-977-7336	Sequence 7336, Ap
39	76.5	9.5	154	15	US-10-335-977-7337	Sequence 7337, Ap
40	76.5	9.5	532	16	US-10-437-963-115812	Sequence 115812, A
41	76	9.4	402	15	US-10-282-122A-53348	Sequence 53348, A
42	75	9.3	206	15	US-10-335-977-8955	Sequence 8955, Ap
43	75	9.3	209	15	US-10-335-977-8956	Sequence 8956, Ap
44	75	9.3	815	15	US-10-282-122A-52922	Sequence 52922, A
45	74.5	9.2	499	15	US-10-424-599-248456	Sequence 248456, A

ALIGNMENTS

RESULT 1
US-09-738-626-6640
; Sequence 6640, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6640
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6640
Query Match 27.0%; Score 218; DB 9; Length 189;

```
Best Local Similarity 33.7%; Pred. No. 1.6e-16;
Matches 57; Conservative 28; Mismatches 60; Indels 24; Gaps 4;

QY 1 MLLPDWKIRK-----EILIEPSESLOPAGYDLVRG--EAFVKG-----40
Db 1 MLLSDRIRKSIDAGDLGIEPDAELIQPSSVDVRMDRYFRVFNNSKYTHIDPKLNQDEL 60

QY 41 --LIDVEGKVIIPREYALITLERIKLPDVMGDMKIRSLAREGVI--GSPAWDP 96
Db 61 TSLVEVDEGEGVLPHGPFVLASTLEKFTLPAHLAGLEGKSSGLRGLLTHSTAGFIDP 120

QY 97 GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQGS 145
Db 121 GFSGYITLELSNVANLPITLWFGMKVQGLALFQMSSPAETPYGSGKLG 169
```

```
RESULT 2
US-10-156-761-11998
; Sequence 11998, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11998
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11998
```

```
Query Match 26.1%; Score 210; DB 14; Length 191;
Best Local Similarity 31.4%; Pred. No. 1.3e-15;
Matches 53; Conservative 32; Mismatches 60; Indels 24; Gaps 3;

QY 1 MLLPDWKIRKEI-----LIEPFSESLOPAGYDLVRG-----EAFVK 38
Db 1 VLLSKDIAEIDAGRVIDPDYDESMVQSSIDVRLDRYFRVFNHRYPHIDPSVEQADL 60

QY 39 GKLIDVEGKVIIPREYALITLERIKLPDVMGDMKIRSLAREGVI--GSPAWDP 96
Db 61 TRLVEPEGDEPFIHPGEPVLASTVEVISLPDLSRLLEGKSSGLRGLVTHSTAGFIDP 120

QY 97 GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQGS 145
Db 121 GFSGHVTLSELNLATLPIKILWFGMKIGQLCMFRLSPAEFPYGSRDYGS 169
```

```
RESULT 3
US-10-437-963-116832
; Sequence 116832, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
```

```
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 116832
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(195)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_20296C.1.pep
US-10-437-963-116832
```

```
Query Match 18.9%; Score 152; DB 16; Length 195;
Best Local Similarity 35.0%; Pred. No. 6.5e-09;
Matches 42; Conservative 18; Mismatches 56; Indels 4; Gaps 3;

QY 22 LOPAGYDLRVGREAFVVGKLLIDVEKEGKVITPPREYALITLERIKLPDDVMGDMKIRSS 81
Db 54 LSGPEQVSAQLEKIMSDEIILAEGEA--FFLHPGELALATTLESVKLPDNIWGLDERS 112

QY 82 LAREGVIGSFA--WVDP--GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY 138
Db 113 LARLGLMVHTAHRIDPXGWHGKIVLEFFNAGKULPAILRLNNAIGALSFEVLGAAAKPY 172
```

```
RESULT 4
US-10-289-762-412
; Sequence 412, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 412
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-412
```

```
Query Match 17.1%; Score 138; DB 15; Length 190;
Best Local Similarity 24.4%; Pred. No. 2.6e-07;
Matches 40; Conservative 31; Mismatches 59; Indels 34; Gaps 5;

QY 13 LIEPP-----SBES-----LQAGYDLRVGRE-----AFVKGKLI 42
Db 18 MIHPVQGVQVNVNBEETGEKLISYGLSSYGYDLRLSREFKVFNTVNVSVVDPKCFTEIDFI 77

QY 43 DVEKEGKVIIPREYALITLERIKLPDVMGDMKIRSLAREGVIGSFAMVDPGWDGML 102
Db 78 SI-TDDVCIVFPNSPALARSVEYFRIPRNVLTMCIGKSTYARCGIIVNVTPPEPEWEGHV 136

QY 103 TMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY---RGNVQ 143
Db 137 TIBISNTTLPKAIYANEGIAQVLFFESSTCTCEVSADRKGYIQ 180
```

```
RESULT 5
US-09-975-719-247
; Sequence 247, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
```

```

; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 247
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-975-719-247

```

Query Match	16.7%	Score 134.5	DB 10	Length 177
Best Local Similarity	26.5%	Pred. No. 5.9e-07		
Matches	41	Conservative 31	Mismatches 64	Indels 19
				Gaps 4
Qy	8	IRKEILTEPSEESL---	QPAGYDLRVG-----	REAFVK---GKLIDVEKGG 49
Db	12	VRDGKLTKLHAHRELTTP	EGVGFDLRLAGLSRLTVGGGS	LRSETRTPASEWLADPDC 71
Qy	50	WVTPPRYVALITLERTKLP	DDVNGDMKIRSSLAREGVIGS	FAMVDPCWDGNLTMLYNA 109
Db	72	FVMEPGKTYLASTWEEFD	LDLALFPFSTLFRSGITFSSSV	LPFGYVGPMTFALTNN 131
Qy	110	SNPEVELRYGERFVOIAF	IRLEGPARPNRYNGOQ 144	
Db	132	HSAPAFETQIGARFAHVIF	QAVSGDI -GRYKGOWOG 165	

```

RESULT 6
US-10-472-928-4752
; Sequence 4752, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4752
; LENGTH: 147
; TYPE: PR1
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: deoxyuridine 5'triphosphate nucleotidohydrolase, putative
; OTHER INFORMATION: Cellular location: cytoplasm
; OTHER INFORMATION: Similar to strain R6 sequence 15902067 (2.E-81)
US-10-472-928-4752

```

Query Match	15.0%;	Score 121;	DB 17;	Length 147;
Best Local Similarity	28.1%;	Pred. No. 1.6e-05;		
Matches	45;	Conservative 23;	Mismatches 56;	Indels 36; Gaps 5;
Qy	7	KIRKEIIIEPSESLOP-----	AGYDLRVG-REAFVKGKLDVKEKGKVVIPREYA	58
		: : : : : : : : :	: : : : : : : : :	
Db	2	KIRGFELVSSFTDENLLPKRETAHAAGYDLKVAVRTVAFGEI	VLVPTGVKAYMQPTEVL	61
Qy	59	LILTLERIKLPDDVWGMKIRSLARE-----	GVIGSFAMVDPCWGNLTLMLYNA	109
		: : : : : : : : :	: : : : : : : : :	
Db	62	YLVD-----	RSSNPRKKGLVLINSVGVDGDDYXGNFGNEGHIFAQMKNI	105
Qy	110	SNPELVLYRGERFYO--	IAFIRLEGPARNPYRGNYOGST	146

```

Db      106 TDQBVLEVGRIVQAVPATFLIADGDAAGVRTGGFGST 145
:::| | | | | : | | | : | | | : | | |
RESULT 7
US-10-617-320-3527
; Sequence 3527, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTO
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; FILING DATE: 10-Jul-2003
; PRIORITY APPLICATION NUMBER: US/10/617,320
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Denke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3527:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...158
; SEQUENCE DESCRIPTION: SEQ ID NO: 3527:
US-10-617-320-3527

Query Match          15.0%; Score 121; DB 18; Leng
Best Local Similarity 28.1%; Pred. No. 1.8e-05;
Matches 45; Conservative 23; Mismatches 56; Ind

QY      7 KIRKEILIEPPSESRLQP-----AGVDLRVG-REAFVKGLKLI
|||:| |:|:| | | |||:| |:|:| | | | | | | | |
Db      13 KIRGFELVSGTSDENLLPRETAHNAAGYDLKVAVRTVVAPGEIV
|||||:| |:|:| | | |||:| |:|:| | | | | | | | |
QY      59 LILTLERIKLPDDVGMGDMKIRSLARE-----GVIGSFPAW
:| | | | | | | | | | | | | | | | | | | | | |
Db      73 YLYD-----RSSNPKKGVLINSGVIDGDYGY
:| | | | | | | | | | | | | | | | | | | | | |

QY      110 SNEPVELRYGBRFVQ---IAFIRLEGPARNPVRNGYQGST 146
:::| | | | | : | | | : | | | : | | |
Db      117 TDQBVLEVGRIVQAVPATFLIADGDAAGVRTGGFGST 156
:::| | | | | : | | | : | | | : | | |

```



```
; FILE REFERENCE: BIOBANK 012A
; CURRENT APPLICATION NUMBER: US/10/733,878
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 60/432699
; PRIOR FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/485027
; PRIOR FILING DATE: 2003-07-03
; NUMBER OF SEQ ID NOS: 535
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 409
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-733-878-409

Query Match      13.8%; Score 111; DB 16; Length 164;
Best Local Similarity 30.7%; Pred. No. 0.00027;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

QY 21 SIQAGYDLRVGREAFVKGKLDVKEGKVIPPPEYALILTLERIKLPDDVGMGMKIRS 80
Db 42 SARAAGYDLYSAYD-----YTIIPPEKAVVKTDIQIALPSGCGYGRVAPRS 86

QY 81 SIAREGVIGSFAMV-DPGWGDGNTLMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 87 GLAAKHFDVAGVIDEDYRGNGVGVLFNFGKEKEFKGDRIAQLICERIFYP 140

RESULT 11
US-10-136-728-10
; Sequence 10, Application US/10136728
; Publication No. US20030236188A1
; GENERAL INFORMATION:
; APPLICANT: Spyttek, Kimberly A.
; APPLICANT: Li, Li
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Stone, David J.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Anderson, David W.
; APPLICANT: Patturajan, Meera
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Smithson, Glenda
; APPLICANT: MacDougall, John R.
; APPLICANT: Mezes, Peter S.
; APPLICANT: Perman, John A.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: No. US20030236188A1el Human Proteins, Polynucleotides Encoding Th
; FILE REFERENCE: 21402-347 D (Cura 647 Other)
; CURRENT APPLICATION NUMBER: US/10/136,728
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 60/288,395
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/289,087
; PRIOR FILING DATE: 2001-05-07
; PRIOR APPLICATION NUMBER: 60/289,619
; PRIOR FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: 60/289,818
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 60/289,817
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 60/290,194
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/290,753
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/291,189
; PRIOR FILING DATE: 2001-05-15

; FILE REFERENCE: BIOBANK 012A
; CURRENT APPLICATION NUMBER: US/10/733,878
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 60/432699
; PRIOR FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/485027
; PRIOR FILING DATE: 2003-07-03
; NUMBER OF SEQ ID NOS: 535
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 409
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-733-878-409

Query Match      13.8%; Score 111; DB 16; Length 164;
Best Local Similarity 30.7%; Pred. No. 0.00027;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

QY 21 SIQAGYDLRVGREAFVKGKLDVKEGKVIPPPEYALILTLERIKLPDDVGMGMKIRS 80
Db 42 SARAAGYDLYSAYD-----YTIIPPEKAVVKTDIQIALPSGCGYGRVAPRS 86

QY 81 SIAREGVIGSFAMV-DPGWGDGNTLMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 87 GLAAKHFDVAGVIDEDYRGNGVGVLFNFGKEKEFKGDRIAQLICERIFYP 140

RESULT 12
US-10-336-472-30
; Sequence 30, Application US/10336472
; Publication No. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderma, Steven K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spyttek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
```

```
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 30
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-30

Query Match 13.8%; Score 111; DB 15; Length 229;
Best Local Similarity 30.7%; Pred. No. 0.00042;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

QY 21 SLQAGYDLVRGREAFAVKGKIDVEKEGKVIPPPEYALILTLERIKLPDDVMGDMKIRS 80
DB 107 SARAAGYDLYSAYD-----YTIIPMEKAVVKTIDQIALPSCGCGRVAPRS 151

QY 81 SLAREGVITGSPAWV-DPGWDGNLTLMLYNASNEPVLELYGRFVQIAIRLEGP 133
DB 152 GLAAKHFDVGVAGVIDEDYRGNVGVLFNFGKEFEVKKGDRIQAOLICERIFYP 205

RESULT 13
US-10-136-728-8
; Sequence 8, Application US/10136728
; Publication No. US20030236188A1
; GENERAL INFORMATION:
; APPLICANT: Spytex, Kimberly A.
; APPLICANT: Li, Li
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Stone, David J.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Anderson, David W.
; APPLICANT: Patturajan, Meera
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Smithson, Glenda
; APPLICANT: MacDougall, John R.
; APPLICANT: Mezes, Peter S.
; APPLICANT: Perman, John A.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: No. US20030236188A1el Human Proteins, Polynucleotides Encoding TH
; FILE REFERENCE: 21402-347 D (Cura 647 Other)
; CURRENT APPLICATION NUMBER: US/10/136,728
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 60/288,395
```

```
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/289,087
; PRIOR FILING DATE: 2001-05-07
; PRIOR APPLICATION NUMBER: 60/289,619
; PRIOR FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: 60/289,818
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 60/289,817
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 60/290,194
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/290,753
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/291,189
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/292,374
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/293,107
; PRIOR FILING DATE: 2001-05-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SEQ ID NO 8
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-136-728-8

Query Match 13.8%; Score 111; DB 15; Length 252;
Best Local Similarity 30.7%; Pred. No. 0.00048;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

QY 21 SLQAGYDLVRGREAFAVKGKIDVEKEGKVIPPPEYALILTLERIKLPDDVMGDMKIRS 80
DB 130 SARAAGYDLYSAYD-----YTIIPMEKAVVKTIDQIALPSCGCGRVAPRS 174

QY 81 SLAREGVITGSPAWV-DPGWDGNLTLMLYNASNEPVLELYGRFVQIAIRLEGP 133
DB 175 GLAAKHFDVGVAGVIDEDYRGNVGVLFNFGKEFEVKKGDRIQAOLICERIFYP 228

RESULT 14
US-10-336-472-28
; Sequence 28, Application US/10336472
; Publication No. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderna, Steven K.
```

```
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Sheno, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Wolenc, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,691
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 28
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-28

Query Match      13.8%; Score 111; DB 15; Length 252;
Best Local Similarity 30.7%; Pred. No. 0.00048;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

Qy 21 SLOPAGYDLRVGREAFVKGKLIIDVEKEGVVIPPREYALILTLERIKLPDDVMGDMKIRS 80
Db 130 SARAAGYDLYSAYD-----YTIPPEKAVVKTDTQIALPSCGYGRVAPRS 174

Qy 81 SLAREGVIGSFANV-DPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 175 GLAAKHFDVAGVIDEDYRGNGVWLFNFGKKEFEVKGKGDRIAQLICERIFYP 228

Search completed: July 2, 2005, 12:34:05
Job time : 31.9767 secs
```

```
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Sheno, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Wolenc, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,691
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 28
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-28

Query Match      13.8%; Score 111; DB 15; Length 252;
Best Local Similarity 30.7%; Pred. No. 0.00048;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

Qy 21 SLOPAGYDLRVGREAFVKGKLIIDVEKEGVVIPPREYALILTLERIKLPDDVMGDMKIRS 80
Db 130 SARAAGYDLYSAYD-----YTIPPEKAVVKTDTQIALPSCGYGRVAPRS 174

Qy 81 SLAREGVIGSFANV-DPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 175 GLAAKHFDVAGVIDEDYRGNGVWLFNFGKKEFEVKGKGDRIAQLICERIFYP 228

RESULT 15
US-10-408-765A-2116
; Sequence 2116, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
```

THIS PAGE LEFT BLANK

Result No.	Query			Description		
	Score	Match	Length	ID		
1	806	100.0	156	2	AAW72847	Polymerase
2	703	87.2	154	4	AAW96424	Putative
3	242	30.0	173	2	AAW51078	Adk16665
4	227.5	28.2	198	8	ADK16665	Nanoarcha
5	220	27.3	193	7	ADM25685	Hyperther
6	218	27.0	189	4	AAW92886	C glutami
7	196	24.3	160	4	AAW80139	Corynebac
8	191.5	23.8	193	5	ABP66173	Bifidobac
9	185.5	23.0	197	6	ADA33378	Acinetoba
10	181	22.5	188	2	AAW89818	Protein e
11	181	22.5	190	2	AAW89897	Antigen 3
12	172	21.3	202	7	ABO68916	Pseudomon
13	170.5	21.2	194	6	ABM68738	Photornab
14	168.5	20.9	192	8	ADL05995	M. catar
15	168.5	20.9	204	7	ADF04778	Bacterial
16	164	20.3	154	7	ADM26960	Hyperther
17	164	20.3	217	7	ABO61782	Klebsiell
18	156	19.4	188	6	ABP78401	N. gonorr
19	155.5	19.3	139	5	ABU52178	Helicobac
20	154.5	19.2	191	8	ADO57546	Actinobac
21	149.5	18.5	190	2	AAV37187	Protein i
22	143.5	17.8	1881	8	ADK15470	Taro bac
23	138	17.1	190	2	AAV34994	Protein i
24	134.5	16.7	177	2	AAV29169	Amino aci
25	121	15.0	147	6	ABU02796	S. pneumo

PCR. p45 and p50 (see AAW7284

This is the amino acid sequence of the P45 component of the polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. The sequence is predicted from a DNA sequence (see AAV63860) obtained from genomic DNA by PCR. P45 and P50 (see AAW72844) are the predominant components of PEF, which acts to enhance the activity of *P. furiosus* DNA polymerase. P45 functions as a dUTPase, and can be used to enhance nucleic acid


```

Db 123 YVDGWRGLTLELVNFMPPVKLRPGNRVQIVFHLSSPAERYAESSCKYHGDERS 182
Qy 150 FSK 152
Db 183 PSK 185

RESULT 6
AAG92886
ID AAG92886 standard; protein; 189 AA.
XX
AC AAG92886;
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum protein fragment SEQ ID NO: 6640.
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-00127688.
XX
XX 16-DEC-1999; 99JP-00377484.
XX
XX 07-APR-2000; 2000JP-00159162.
XX
XX 03-AUG-2000; 2000JP-00280988.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
XX N-PSDB; AAH68105.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analyzing
XX expression profile or pattern of a gene and identifying homologous gene.
XX
XX Claim 17; SEQ ID NO 6640; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and analysing
XX the expression profile or expression pattern of a gene derived from
XX Coryneform bacterium, and identifying a homologue of a gene derived from
XX coryneform bacterium. Coryneform bacteria are useful for producing amino
XX acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described in the
XX exemplification of the invention. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from the European Patent Office
XX
XX Sequence 189 AA;

Query Match 27.0%; Score 218; DB 4; Length 189;
Best Local Similarity 33.7%; Pred. NO. 1.3e-17;
Matches 57; Conservative 28; Mismatches 60; Indels 24; Gaps 4;

Qy 1 MLLPDWKIRK-----EILIEPSEESLQAGYDLRVR--EAFVKGK----- 40
Db 1 MLLSDRDIRKSIDAGDLGIEPFDALIQPSSVDVMDRYFRVFNNSKYTHIDPKLNQDEL 60
Qy 41 --LIDVEKGKVIIPREYALITLIERIKLPDDVMGDKIRGSLAREGVI--GSFAWVDP 96
Db 61 TSLVEDEGGEFVLPGEFVLASTLEKFTLPAHLAAGRLGKSLGLRLTHSTAGFIDP 120
```

```

Qy 97 GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYOQS 145
Db 121 GFSGYITLLESNVANLPITLWPMKMGVQALFQMSSPAETPYGSGKLG 169

RESULT 7
AAB80139
ID AAB80139 standard; protein; 160 AA.
XX
AC AAB80139;
XX
XX 30-APR-2001 (first entry)
XX
XX Corynebacterium glutamicum MP protein sequence SEQ ID NO:1012.
XX
XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;
XX fine chemical production; microorganism; organic acid; nucleoside;
XX nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
XX lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
XX carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
XX
XX OS
XX OS Corynebacterium glutamicum.
XX
XX WO200100843-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-IB000923.
XX
XX 25-JUN-1999; 99US-0141031P.
XX 01-JUL-1999; 99DE-01030476.
XX 02-JUL-1999; 99US-0142101P.
XX 08-JUL-1999; 99DE-01031415.
XX 08-JUL-1999; 99DE-01031418.
XX 08-JUL-1999; 99DE-01031419.
XX 08-JUL-1999; 99DE-01031420.
XX 08-JUL-1999; 99DE-01031424.
XX 08-JUL-1999; 99DE-01031428.
XX 08-JUL-1999; 99DE-01031434.
XX 08-JUL-1999; 99DE-01031435.
XX 08-JUL-1999; 99DE-01031443.
XX 08-JUL-1999; 99DE-01031453.
XX 08-JUL-1999; 99DE-01031457.
XX 08-JUL-1999; 99DE-01031465.
XX 08-JUL-1999; 99DE-01031478.
XX 08-JUL-1999; 99DE-01031510.
XX 08-JUL-1999; 99DE-01031541.
XX 08-JUL-1999; 99DE-01031573.
XX 08-JUL-1999; 99DE-01031592.
XX 08-JUL-1999; 99DE-01031632.
XX 08-JUL-1999; 99DE-01031634.
XX 08-JUL-1999; 99DE-01031636.
XX 08-JUL-1999; 99DE-01032125.
XX 08-JUL-1999; 99DE-01032126.
XX 08-JUL-1999; 99DE-01032130.
XX 08-JUL-1999; 99DE-01032186.
XX 08-JUL-1999; 99DE-01032206.
XX 08-JUL-1999; 99DE-01032227.
XX 08-JUL-1999; 99DE-01032228.
XX 08-JUL-1999; 99DE-01032229.
XX 08-JUL-1999; 99DE-01032230.
XX 08-JUL-1999; 99DE-01032922.
XX 08-JUL-1999; 99DE-01032926.
XX 08-JUL-1999; 99DE-01032928.
XX 08-JUL-1999; 99DE-01033004.
XX 08-JUL-1999; 99DE-01033005.
XX 08-JUL-1999; 99DE-01033006.
XX 08-JUL-1999; 99US-0148613P.
XX 08-JUL-1999; 99DE-01040764.
XX 08-JUL-1999; 99DE-01040765.
XX 08-JUL-1999; 99DE-01040766.
XX 08-JUL-1999; 99DE-01040832.
```



```

PR 31-AUG-1999; 99DE-01041378.
PR 31-AUG-1999; 99DE-01041379.
PR 31-AUG-1999; 99DE-01041380.
PR 31-AUG-1999; 99DE-01041394.
PR 31-AUG-1999; 99DE-01041396.
PR 03-SEP-1999; 99DE-01042076.
PR 03-SEP-1999; 99DE-01042077.
PR 03-SEP-1999; 99DE-01042079.
PR 03-SEP-1999; 99DE-01042086.
PR 03-SEP-1999; 99DE-01042087.
PR 03-SEP-1999; 99DE-01042088.
PR 03-SEP-1999; 99DE-01042095.
PR 03-SEP-1999; 99DE-01042124.
PR 03-SEP-1999; 99DE-01042129.
PR 03-MAR-2000; 2000US-0187970P.
XX
PA (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-137957/14.
XX N-PSDB; AAF72258.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic pathway
PT proteins, useful for producing fine chemicals in microorganisms,
PT including organic acids, nonproteinogenic amino acids, and purine and
PT pyrimidine bases.
XX
XX Claim 20; Page 1554; 1737pp; English.
XX
XX AAF71753 to AAF72130 encode the Corynebacterium glutamicum metabolic
XX pathway (MP) proteins given in AAB79634 to AAB80211. The C. Glutamicum MP
XX nucleic acids are useful for the production of fine chemicals in
XX microorganisms, including organic acids, nonproteinogenic amino acids,
XX purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated
XX and unsaturated fatty acids, diols, carbohydrates, aromatic compounds,
XX vitamins, cofactors, polyketides and enzymes
XX
XX Sequence 160 AA;
XX
XX Query Match 24.3%; Score 196; DB 4; Length 160;
XX Best Local Similarity 32.9%; Pred. No. 4.7e-15;
XX Matches 52; Conservative 29; Mismatches 53; Indels 24; Gaps 4;
XX
QY 1 MLLPDWKIRK-----RLIEPFSESLQAGYDLVRG--EAFVKGK-----40
Db 1 VLLSDRDIRKSIDAGDLGIEPPDAELIQSSVDVRMDRVFRVFNNSKYTHIDPKLNQDEL 60
QY 41 --LIDVEKEGKVIPPREYALILTLERIKLPDDVMDGMIKIRSLAREGVI--GSFAWVDP 96
Db 61 TSLVEVEDEGEGFVLLHPGEFVLAFTLEKFTLPAHLAQLRLEKSLGLRLGLLTHSTAGFIDP 120
QY 97 GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPA 134
Db 121 GFSGYITLSENVNLPITLWPGMKVQGLALFQMSSPA 158
XX
XX RESULT 8
XX ID ABP66173
XX ID ABP66173 standard; protein; 193 AA.
XX
XX AC ABP66173;
XX
XX 19-NOV-2002 (first entry)
XX
XX Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:917.
XX
XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
XX anti-diarrheic; antibacterial; inhibitor of Salmonella; detection;
XX identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
XX rotavirus; food composition; pharmaceutical composition.
XX
XX Bifidobacterium longum.

```

```

XX EPI227152-AL.
XX
XX 31-JUL-2002.
XX
XX 30-JAN-2001; 2001EP-00102050.
XX
XX 30-JAN-2001; 2001EP-00102050.
XX (NEST ) SOC PROD NESTLE SA.
XX
XX WPI; 2002-668397/72.
XX
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
XX a probe or primer for detecting and/or identifying Bifidobacterium longum
XX in a biological sample.
XX
XX Claim 3; SEQ ID NO 917; 80pp; English.
XX
XX The present invention describes a polynucleotide (I) comprising a
XX sequence of a Bifidobacterium genome selected from the nucleotide
XX sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
XX least 90% identity or which hybridises with the sequences given in
XX ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
XX fusion protein, comprising a sequence selected from 1097 sequences given
XX in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
XX heterologous polypeptide. (I) has anti-diarrheic and antibacterial
XX activities, and can be used as an inhibitor of Salmonella. (I) (which is
XX a probe) is useful for the detection and/or identification of
XX Bifidobacterium longum in a biological sample. A carrier containing the
XX lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-3618) can be
XX used for preventing and/or treating diarrhoea brought about by pathogenic
XX bacteria and/or rotavirus. The carrier is a food composition selected
XX from milk, yogurt, curd, cheese, fermented milks, milk based fermented
XX products, ice-creams, fermented cereal based products, milk based
XX powders, infant formula, pet food or a pharmaceutical composition
XX selected from tablets, liquid bacterial suspensions, dried oral
XX supplement, wet oral supplement, dry tube feeding or wet tube feeding.
XX (I) is useful in DNA arrays or chips to carry out analysis of the
XX expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
XX Bifidobacterium related nucleotide sequences given in the Sequence
XX Listing from the present invention but not mentioned further within the
XX specification. N.B. The sequence data for this patent is not represented
XX in the printed specification but is based on sequence information
XX supplied by the European Patent Office
XX
XX Sequence 193 AA;
XX
XX Query Match 23.8%; Score 191.5; DB 5; Length 193;
XX Best Local Similarity 31.2%; Pred. No. 2.2e-14;
XX Matches 54; Conservative 33; Mismatches 57; Indels 29; Gaps 5;
XX
QY 1 MLLPDWKI-----RKELIEPFSESLQAGYDLVRG-----EAFVKGKL 41
Db 1 MLLSDRDLAAQSAGHISLDPTPEMVQPSADIVLRDYLFRLENNHAYTVYDPAENQAL 60
QY 42 ---IDVEKEGKVIPPREYALILTLERIKLPDDVMDGMIKIRSLAREGVI--GSFAWVDP 96
Db 61 TEQFEVAPDEPWLHPGEFALGSTWEYVKLDPSIAARLEKSLGLRLGLLTHSTAGFIDP 120
QY 97 GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY-----RGNYYQG 144
Db 121 GFEGHITLSENVSTLPVKLWPGMKIQMCFQLQSSPAEHPYGSKGTGSHYQG 173
XX
XX RESULT 9
XX ADA33378
XX ID ADA33378 standard; protein; 197 AA.
XX
XX AC ADA33378;
XX
XX DT 20-NOV-2003 (first entry)
XX
XX

```

```
DE Acinetobacter baumannii protein #539.
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
FN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
WPI; 2003-576092/54.
DR N-PSDB; ADA29252.
XX
New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 4665; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
SQ Sequence 197 AA;
Query Match 23.0%; Score 185.5; DB 6; Length 197;
Best Local Similarity 27.4%; Pred. No. 1.2e-13;
Matches 48; Conservative 31; Mismatches 63; Indels 33; Gaps 4;
QY 10 KEILTEPPEESLQ-----PAGYDLRVGRE-----AFVKGK 40
DB 23 KHGMTEPYAENQVRFDKNGEKLISYGVSSGYDVRCAREFKVTNNVHSAIVDPKNFDEKS 82
QY 41 LIDVEKEGKVVPPREYALILTLERIKLPPDDVGMKIRSSLSAREGVIGSPFAWVDPGWDG 100
DB 83 FIDIESD-VCIIPPNFSFALARTIEYFRPNVLTVCGLGKSTYARCGIIVNVTPLSPFMEG 141
QY 101 NUTLMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYR---GNYQGSTRLAFSK 152
DB 142 HVTLEFSNTNLPARIYAGEVQAQMLPFESDEVCSYKDRGKYQGQGTGVLTPK 196
RESULT 10
AAW89818
ID AAW89818 standard; protein; 188 AA.
XX
AC AAW89818;
XX
DT 20-MAR-2003 (revised)
DT 18-FEB-1999 (first entry)
XX
DE Protein encoded by clone B17 ORF4.
XX
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX
OS Helicobacter pylori.
```

```
XX WO9849314-A2.
XX
PD 05-NOV-1998.
XX
PF 25-APR-1998; 98WO-US008487.
XX
PR 25-APR-1997; 97US-0045107P.
PR 14-OCT-1997; 97US-0061958P.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Chow TP, Fry KE, Lim MY, Mcatee CP;
XX
WPI; 1999-009433/01.
DR N-PSDB; AAV90555.
XX
New Helicobacter pylori antigens and related nucleic acid sequences -
PT useful in serological diagnosis and protective vaccines, providing long-
PT lasting immune response.
XX
PS Claim 15; Page 104-105; 402pp; English.
XX
CC The present sequence represents a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma). (Updated on 20-MAR-2003
CC to correct PF field.)
XX
SQ Sequence 188 AA;
Query Match 22.5%; Score 181; DB 2; Length 188;
Best Local Similarity 28.0%; Pred. No. 3.9e-13;
Matches 47; Conservative 29; Mismatches 64; Indels 28; Gaps 4;
QY 13 LIETPFSE-----SLQPAGYDLRVGRE-----AFVKGK-----LIDVEKE 47
DB 18 MISPFCEKQVGNVTISYGLSSGYDIRVSEFMLFDKNKALIDPKNFDPNNAKIDASKE 77
QY 48 GKVVIPPREYALILTLERIKLPPDDVGMKIRSSLSAREGVIGSPFAWVDPGWDGNTLMY 107
DB 78 GYFILPANAFALAHITIEYFKMPKDTLAI CLGKSTYARCGIIVNVTPEPEFEGYITIEIS 137
QY 108 NASNEPVELRYGERFVQIAFIRLEGPARNPYR---GNYQGSTRLAFSK 152
DB 138 NTTNLPKAVYANEGIAQVVFLOQDEMCESQSYKDRGKYQGQGTGVLTPK 185
RESULT 11
AAW89897
ID AAW89897 standard; protein; 190 AA.
XX
AC AAW89897;
XX
DT 20-MAR-2003 (revised)
DT 18-FEB-1999 (first entry)
XX
DE Antigen 3 from cluster 28c.
XX
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX
OS Helicobacter pylori.
XX
FN WO9849314-A2.
XX
```



```
PR 07-FEB-2001; 2001FR-00001659.
XX (INSP ) INST PASTEUR.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
XX WPI; 2003-148459/14.
DR
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX Claim 2; SEQ ID NO 1835; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 194 AA;
Query Match 21.2%; Score 170.5; DB 6; Length 194;
Best Local Similarity 29.1%; Pred. NO. 7.7e-12;
Matches 51; Conservative 28; Mismatches 61; Indels 35; Gaps 6;
QY 6 WKIRKELIEPSE-ESLQPAGYDLRVGRE-----AFVK-----CKLIDVEKE 47
DB 11 WLDEGKLVIAAPPIERINGATADVRLGNQFRVFCGHTAAYIDLSPKDEVSAAALDRVMS 70
QY 48 GKWVIP-----PREVALILTLEKILPDDVMGDMKIRSLAREGVIGSPA--WVDPG 97
DB 71 DEIILPDDDEVFLHPCGELALVLTLSVTLPDDLGVGLDGRSSLARLGLMVHTAHRIDPG 130
QY 98 WDNNTLMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYR-----GNTQGS 145
DB 131 WHGQIVLEFYNGKPLALRPGKVGICALSFEPILSGSADRPYNNRQDAKYKNOOGA 185
RESULT 14
ADL05995
ID ADL05995 standard; protein; 192 AA.
AC ADL05995;
XX
XX 06-MAY-2004 (first entry)
DT
DE M. catarrhalis protein #1761.
XX
XX Moraxella catarrhalis; infection.
XX
XX Moraxella catarrhalis.
OS
XX
XX US6673910-B1.
PN
XX
XX 06-JAN-2004.
PD
XX
XX 07-FEB-2001; 2000US-00540236.
XX
XX 08-APR-1999; 99US-0128416P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;
XX
XX WPI; 2004-178127/17.
DR
XX
XX New nucleic acid encoding a Moraxella catarrhalis polypeptide, useful for
PT preparing a composition for diagnosing, preventing or treating infection
PT caused by Moraxella catarrhalis.
XX
XX Disclosure; SEQ ID NO 3681; 429pp; English.
XX
XX The invention relates to an isolated nucleic acid encoding an Moraxella
CC catarrhalis polypeptide. The nucleic acid is useful for preparing a
CC composition for diagnosing, preventing or treating infection caused by
CC Moraxella catarrhalis. The present sequence represents the amino acid
CC sequence of a M. catarrhalis protein.
XX
SQ Sequence 192 AA;
Query Match 20.9%; Score 168.5; DB 8; Length 192;
Best Local Similarity 28.2%; Pred. NO. 1.3e-11;
Matches 49; Conservative 30; Mismatches 60; Indels 35; Gaps 6;
QY 13 LIEPSEESLQ-----PAGYDLRVGRE-----AFVKGK-----LID 43
DB 19 MIEPEPQVRYNAQGQKLVSYGTSSYGVYRCANEFKVFTNVHSAIVDPKDFDRSFID 78
QY 44 VEKEGVVPPREVALILTLEKILPDDVMGDMKIRSLAREGVIGSPA--WVDPG 103
DB 79 IIGD-ECIIPNSFALARTVEYFRIPRDVLTICLGKSTYARCGIIVNVTPLEPEWEHVT 137
QY 104 LMLYNASNEPVELRYGERFVQIAFIRLEGP--ARNPYR---GNYOGSTRLAFSK 152
DB 138 LEFSNTTLPAIYAGEGVAQMLFFQSDADDVCCISYKDRGKYGGQGTGVLTPK 191
RESULT 15
ADF04778
ID ADF04778 standard; protein; 204 AA.
XX
XX ADF04778;
XX
XX 12-FEB-2004 (first entry)
DT
XX
DE Bacterial polypeptide #891.
XX
XX Proteus mirabilis infection; bacterial infection; antibacterial;
KW immunostimulant.
XX
XX Proteus mirabilis.
OS
XX
XX US6605709-B1.
PN
XX
XX 12-AUG-2003.
PD
XX
XX 05-APR-2000; 2000US-00543681.
PF
XX
XX 09-APR-1999; 99US-0128706P.
PR
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;
XX
XX WPI; 2003-895291/82.
DR
XX
XX N-PSDB; ADF00606.
XX
```


THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 11:27:13 ; Search time 80.7442 Seconds
(without alignments)
2555.824 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 1998
Sequence: 1 MLHHVKLIYATSKRLVGKK.....MKKRELABRIWDEIKXLS 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1933.5	91.8	402	Q8X250	Q8X250 pyrococcus
2	1667.5	83.5	401	Q59114	Q59114 pyrococcus
3	1649.5	82.6	401	Q9V0S3	Q9V0S3 pyrococcus
4	781.5	39.1	466	Q8PUP9	Q8PUP9 methanosarc
5	759.5	38.0	431	Q8TEA0	Q8TEA0 methanosarc
6	735	36.8	404	Q28628	Q28628 archaeoglob
7	731	36.6	382	Q9HH70	Q9HH70 methanobact
8	731	36.6	386	Q27284	Q27284 methanobact
9	648.5	32.5	413	Q980S0	Q980S0 sulfolobus
10	636.5	31.9	412	Q976C4	Q976C4 sulfolobus
11	627.5	31.4	403	1 DFP METJA	Q58323 m coenzyme
12	626	31.3	392	Q9HRS1	Q9HRS1 halobacteri
13	605	30.3	398	Q971C8	Q971C8 clostridium
14	577	28.9	395	Q8XJL0	Q8XJL0 clostridium
15	571	28.6	414	Q67PR7	Q67PR7 symbiobacte
16	563.5	28.2	399	Q74E40	Q74E40 geobacter s
17	557	27.9	398	Q8R9S8	Q8R9S8 thermoanaer
18	555	27.8	396	Q6LWU8	Q6LWU8 methanococc
19	543.5	27.2	390	1 CABG BORBU	Q51752 b coenzyme
20	542	27.1	401	Q819T8	Q819T8 bacillus ce
21	542	27.1	401	Q6HEU6	Q6HEU6 bacillus th
22	541	27.1	401	Q81WG9	Q81WG9 bacillus an
23	540.5	27.1	397	Q895Q4	Q895Q4 clostridium
24	540	27.0	401	Q636F7	Q636F7 bacillus ce
25	537	26.9	401	Q732K1	Q732K1 bacillus ce
26	535.5	26.8	388	Q66997	Q66997 aquifex aeo
27	535	26.8	437	Q9YAI0	Q9YAI0 aeropyrum p
28	533	26.7	427	Q8CK27	Q8CK27 streptomyce
29	529	26.5	400	Q8AB12	Q8AB12 bacteroides
30	527	26.4	405	Q827Q1	Q827Q1 streptomyce
31	510	25.5	353	2 Q652W1	Q652W1 borrelia ga

32	509	25.5	376	2	Q8TXG1	Q8txg1 methanopyru
33	507.5	25.4	414	2	Q7MFS9	Q7mp59 vibrio vuln
34	506	25.3	399	2	Q92AI3	Q92ai3 listeria in
35	505	25.3	401	2	Q8ZXM3	Q8zxm3 pyrobaculum
36	504.5	25.3	403	2	Q8DDX8	Q8ddx8 vibrio vuln
37	502.5	25.2	397	2	Q64S04	Q64s04 bacteroides
38	502.5	25.2	403	2	Q6AMM3	Q6amm3 desulforale
39	500.5	25.1	391	2	Q83D56	Q83d56 coxiella bu
40	500.5	25.1	402	2	Q8CXH5	Q8cxh5 oceanobacil
41	498.5	24.9	401	2	Q7P314	Q7p314 fusobacteri
42	495.5	24.8	404	2	Q8RTI5	Q8rti5 fusobacteri
43	495	24.8	399	2	Q71XJ1	Q71xj1 listeria mo
44	494.5	24.7	404	2	Q9K9Y4	Q9k9y4 bacillus ha
45	494	24.7	399	2	Q8Y674	Q8y674 listeria mo

ALIGNMENTS

RESULT 1

Q8X250 PRELIMINARY; PRT; 402 AA.
AC Q8X250: Q7LWV5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE DNA flavoprotein Dfp (Dna/pantothenate metabolism flavoprotein).
GN OrderedLocusNames=PF1440;
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21664347; PubMed=11782527; DOI=10.1073/pnas.0123727999;
RA Hogrefe H.H., Hansen C.J., Scott B.R., Nielson K.B.;
RT "archaeal dUTPase enhances PCR amplifications with archaeal DNA
polymerases by preventing dUTP incorporation.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:596-601(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY066006; AAL47573.1; -
DR EMBL; AS010247; AAL81564.1; -
DR HSSP; P30197; IG63.
DR Pfam; PF04127; DFP; 1.
DR TIGRFAMs; TIGR00521; coaBC dfp; 1.
SQ SEQUENCE 402 AA; 44721 MW; 9C14A470470CB0D8 CRC64;

Query Match 91.8%; Score 1833.5; DB 2; Length 402;
Best Local Similarity 93.3%; Pred. No. 3.3e-96;
Matches 378; Conservative 6; Mismatches 16; Indels 5; Gaps 4;
QY 1 MLHHVKLIYATSKRLVGKKIVXXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKII 59
DB 1 MLHHVKLIYATSKRLVGKKIVLAIPGSIAAECVKLARELIRHGAEVHVMSEAATKII 60
QY 60 HPYAMNLPITGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDDPVPT 119
DB 61 HPYAMFATGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDDPVPT 120
QY 120 TVVTTAFPHIPMIAPAMHETMYRHPDIVRENIERLKKLVGEFTGPRIEGRKAVASIDEI 179
DB 121 TVVTTAFPHIPMIAPAMHETMYRHPDIVRENIERLKKLVGEFTGPRIEGRKAVASIDEI 180
QY 180 VYRVIKLHKKTLEGRKVLVTAGATREYIDPIRFITNASSGKMGVALAEADFRGA-VTL 238
DB 181 VYRVIKLHKKTLEGRKVLVTAGATREYIDPIRFITNASSGKMGVALAEADFRGA-VTL 240

```

QY 239 IRTKGSVKAIRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
DB 241 IRTKGSVKSFVENQI--EVEETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
QY 299 KSGRSITITLVPXNPKIIDRIKEIQPNVFLVGFKAETSKEKLIBEGKQIERAKADLVVG 358
DB 299 KSDRSITITLVP-NPKIIDRIKEIQPNVFLVGFKAETSKEKLIBEGKQIERAKADLVVG 357
QY 359 NTLEAFGSEENQVLIIGRDFTKELPKMKKRELAERIMWDEIEKLS 403
DB 358 NTLEAFGSEENQVLIIGRDFTKELPKMKKRELAERIMWDEIEKLS 402

RESULT 2
O59114 PRELIMINARY; PRT; 401 AA.
AC O59114;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein PH1444.
GN OrderedLocNames=PH1444;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo H. Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Maechi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000006; BAA30551.1; -.
DR PIR; G71018; G71018.
DR HSSP; P30197; 1G63.
DR InterPro; IPR005252; Cons hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 401 AA; 44624 MW; 26CB5F523AE02C8 CRC64;

Query Match 83.5%; Score 1667.5; DB 2; Length 401;
Best Local Similarity 83.4%; Pred. No. 8.4e-87;
Matches 337; Conservative 33; Mismatches 29; Indels 5; Gaps 4;

QY 1 MLHVKLIYATSKRLVGKKIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEATKII 59
DB 1 MLHVKRIYAKSKRLVGKKIVLAIPGSAAEVCKLARELIRHGAHVHVMTPSATKII 60
QY 60 HPYANLPTGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 119
DB 61 HPYAMEFATGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 120
QY 120 TVVTTAFPHIPIMAPAMHETMYRHPIVRENTERLKKLVGEFIPRIEGRKVASIDEI 179
DB 121 TVVTTAFPHIPIMAPAMHESMYKHPIVRENIEKLKLVGEFIPRIEGRKVASIDEI 180
QY 180 YRVVTKLHKHTLEGKRVLTAGATREYIDPFRFITNASSGGMGVALAEAEADFRGA-VTL 238
DB 181 YRVVTKLHKHTLEGKRVLTAGATREYIDPFRFITNASSGGMGVALAEAEADFRGA-VTL 240
QY 239 IRTKGSVKAIRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
DB 241 IRTKGSVKSFVENQI--EVEETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
QY 299 KSGRSITITLVPXNPKIIDRIKEIQPNVFLVGFKAETSKEKLIBEGKQIERAKADLVVG 358
DB 299 KSDRSITITLVP-NPKIIDRIKEIQPNVFLVGFKAETSKEKLIBEGKQIERAKADLVVG 357

```

```

DB 241 IRTKGSVNSFVENQI--QVETVEEMLAAIEKELTEKKYDVVIMAAVSDFKPKIKAEDKI 298
QY 299 KSGRSITITLVPXNPKIIDRIKEIQPNVFLVGFKAETSKEKLIBEGKQIERAKADLVVG 358
DB 299 KSNKSITITLVP-NPKIIDRIKEIQPNVFLVGFKAETTWKLIQAKKQIERAKSDIVIG 357
QY 359 NTLEAFGSEENQVLIIGRDFTKELPKMKKRELAERIMWDEIEKKL 402
DB 358 NTLEAFGSDSKVLIIGRNFEKELPKMKKRELAERIMWDEIEKRL 401

RESULT 3
Q9V0S3 PRELIMINARY; PRT; 401 AA.
AC Q9V0S3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Dfp DNA/pantothenate metabolism flavoprotein.
GN ORFNames=PAB1897;
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL MOL. Microbiol. 47:1495-1512(2003).
DR EMBL; AJ248285; CAB49630.1; -.
DR PIR; E75114; E75114.
DR HSSP; P30197; 1G63.
DR InterPro; IPR005252; Cons hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 401 AA; 44290 MW; 96CCD5A190C1FA5D CRC64;

Query Match 82.6%; Score 1649.5; DB 2; Length 401;
Best Local Similarity 82.9%; Pred. No. 8.8e-86;
Matches 335; Conservative 31; Mismatches 33; Indels 5; Gaps 4;

QY 1 MLHVKLIYATSKRLVGKKIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEATKII 59
DB 1 MLSHKLIYATSKRLVGKKIVLAIPGSAAEVCKLARELIRHGAHVHVMSPATKII 60
QY 60 HPYANLPTGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 119
DB 61 HPYAMEFATGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 120
QY 120 TVVTTAFPHIPIMAPAMHETMYRHPIVRENTERLKKLVGEFIPRIEGRKVASIDEI 179
DB 121 TVVTTAFPHIPIMAPAMHESMYKHPIVRENIEKLKLVGEFIPRIEGRKVASIDEI 180
QY 180 YRVVTKLHKHTLEGKRVLTAGATREYIDPFRFITNASSGGMGVALAEAEADFRGA-VTL 238
DB 181 YRVVTKLHKHTLEGKRVLTAGATREYIDPFRFITNASSGGMGVALAEAEADFRGA-VTL 240
QY 239 IRTKGSVKAIRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
DB 241 IRTKGSVKSFVENQI--EVEETVEEMLGAIEBELTKKYDVVIMAAVSDFKPKIKAEGKI 298
QY 299 KSGRSITITLVPXNPKIIDRIKEIQPNVFLVGFKAETSKEKLIBEGKQIERAKADLVVG 358
DB 299 KSDKSITITLVP-NPKIIRKIKIEIQSDVFLVGFKAETSMKELISEAKKQIESAGSDLVIG 357

```


QY 359 NTLEAFSGSENOVLIGRDTFKLPKMKRELAERIWDIEIKXL 402
DB 358 NTLEAFSGSEKVVIVGKDFKLPKMKRELAERIWDIEIKKI 401

RESULT 4
Q8PUP9

ID Q8PUP9 PRELIMINARY; PRT; 466 AA.
AC Q8PUP9; (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DE DNA/pantothenate metabolism flavoprotein.
GN OrderedLocusNames=NM2283;
OS Methanosarcina mazel (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goel / GoI / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wieser A., Baumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazel: evidence for lateral gene
RT transfer between Bacteria and Archaea";
RL J. Mol. Microbiol. Biotechnol. 4:453-461 (2002).
RL EMBL: AS013470; AM31979.1; -;
DR HSSP; Q9RC23; 1P3Y.
DR InterPro; IPR005252; Cons hyp0521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coabc_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 466 AA; 50459 MW; 7290D4F35FC5AA25 CRC64;

Query Match 39.1%; Score 781.5; DB 2; Length 466;
Best Local Similarity 45.0%; Pred. No. 1.7e-36;
Matches 186; Conservative 70; Mismatches 136; Indels 21; Gaps 13;
QY 2 LHHVKLIYATKSRKLVGKIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKIIH 60
DB 59 IHPTLWIOGSSSLAGKIVLVGTVSGIGAVRVVLSRELIRNGAEVHVMSEAATKIIH 118
QY 61 PYAWNLPNGPVITEITGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTV 120
DB 119 PDALHYATGNPVAELGGRVHEVFCG-FKGRADLLLIAPATANTIGKIAYGIDDTVTS 177
QY 121 VTTAP-PIPIPIAMHETMYRHPVIRENIRLKLGVFEFGPIREGRKVASIDEI 179
DB 178 FATTALGSDVPVIVPAMHESMYRHPAVTENIARLKSWMGIVSGPRFEGVAKIAGNEE 237
QY 180 VYRVIKLHKTKLEGRVLTAGATREYIDPIRITNASSGKMGVALAEADFRGA-VTL 238
DB 238 VLEVERALGDRSLNKVITGSGTASLDPILVITNRSAGKTGRELALEYRGADVTL 297
QY 239 I-RTKGSVKAFIRIKLVKIVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEG 297
DB 298 VHRDRIGFAGIR----EIFAESAAMEAVLSEL-EKGYDALISSAAIADYTAEPSP-K 351
QY 298 IKSGRSITIELVPXNKPIDRIKEIQPNVFLVGFKAET--SKKLIEEGKQIERAKADL 355
DB 352 IKSGGFIKLKPKTR-KLIKECREKYPDELVIIGFKAETGIEKELLRRAAATLEGTELD 410
QY 356 VVGNTL--EAFSGSENOVLIGRDTFKLPKMK--KRELAERIWDIEIKXL 403
DB 411 IANDVKGKGMGTENELLYLGK--KSGPRHVSNGNRKRLAASILBEVSEILN 461

RESULT 5
Q8TRA0

ID Q8TRA0 PRELIMINARY; PRT; 431 AA.
AC Q8TRA0; (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Pantothenate metabolism flavoprotein.
GN Names=dpf; OrderedLocusNames=MA1280;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity";
RL Genome Res. 12:532-542 (2002).
DR EMBL: AE010797; AM04699.1; -;
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coabc_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 431 AA; 46193 MW; 646596DDC9B4C156 CRC64;

Query Match 38.0%; Score 759.5; DB 2; Length 431;
Best Local Similarity 43.5%; Pred. No. 2.8e-35;
Matches 178; Conservative 78; Mismatches 136; Indels 17; Gaps 12;
QY 3 HHVKLIYATKSRKLVGKIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKIIHP 61
DB 27 HPTLWIOGKSSLSKIVLVGTVSGIGAVRVVLSRELIRNGAEVHVMSEAATKIIHP 86
QY 62 YAWNLPNGPVITEITGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTV 121
DB 87 DALQYATGNPVTIELGGRVHEVFCG-FKGRADLLLIAPATANTIGKIACGIDDTVTSF 145
QY 122 VTTAP-PIPIPIAMHETMYRHPVIRENIRLKLGVFEFGPIREGRKVASIDEI 180
DB 146 ATTALGSGVPLVMPVMPAMHESMYRHPAVTENIARLKSWMGIVSGPRFEGVAKIAGNEE 205
QY 181 YRVIKLHKTKLEGRVLTAGATREYIDPIRITNASSGKMGVALAEADFRGA-VTLI 239
DB 206 LVEVERALGDRSLNKVITGSGTASLDPILVITNRSAGKTGRELALEYRGADVTL 265
QY 240 -RTKGSVKAFIRIKLVKIVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEG 298
DB 266 HRDRILGLAGIK----EVFAESAAMEAVLSEL-EKGYDALISSAAIADYTAEPSP-KI 319
QY 299 KSGRSITIELVPXNKPIDRIKEIQPNVFLVGFKAET--SKKLIEEGKQIERAKADLV 356
DB 320 KSGGFEVLKPKTR-KLIKECREKYPDELVIIGFKAETGVRGDRLLRRAAATLGTKLDLI 378
QY 357 VVGNTL--EAFSGSENOVLIGRDTFKLPKMKRELAERIWDIEIKXL 403
DB 379 AANDVKGKGMGTENELLYLGKPRHV-SGNKRKLAACILELTGLN 426

```
RESULT 6
O28628      PRELIMINARY;          PRT;    404 AA.
AC O28628;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE Pantothenate metabolism flavoprotein (Dfp).
GN OrderedLocusNames=AF1645;
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475; DOI=10.1038/37052;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervilange A.R., Graham D.E., Kyprides N.C.,
RA Pleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.J.,
RA Peterson S.N., Reich C.I., McNeil L.K., Badger J.H., Glodek A.,
RA Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,
RA Uitterback T.R., Cotton M.D., Spriggs T., Artiach P., Kaine B.P.,
RA Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,
RA Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,
RA Woese C.R., Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL; AE000989; AAB89597.1; -.
DR PIR; D69455; D69455.
DR TIGR; AF1645; -.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 404 AA; 44885 MW;  E5A2B899C3E6A66F CRC64;

Query Match      36.8%; Score 735; DB 2; Length 404;
Best Local Similarity 45.0%; Pred. No. 6.4e-34;
Matches 182; Conservative 74; Mismatches 122; Indels 26; Gaps 13;
QY 4 HVKLIYATKSRKLVGKVIYXXPGSIALD-VKACEGLIRHGAHVHVMSEAAATKIHPY 62
DB 5 HLERIRGRSRKLERKKIVLGVGTGSIAAVETVKLARELVRRGADVAVMSRAARKIHPY 64
QY 63 ANWLPTGNPVITEGFIHVELAGHENKADLLVCPATANTISKIACGIDDTPTVTVV 122
DB 65 ALEFATGKRVVTEITGSIHVNLLGEY-GDADFLIAPCTANTISKIACGIDDTPTTFA 123
QY 123 TTAF-PHIPMTAPAMHTMYRHPVIRENIRLKLGVFIPRIEGRKAVASIDEIVY 181
DB 124 TTLGSGKPIIIVPAMHEAMNKAIVLENIQRLIDMGIEFVOPRIEGRKAVPSTETCL 183
QY 182 RVIKKLHKHTLGRKRVLTAGATREYIDPPIRITNASSGKMGVALAEADPRGA-VTLIR 240
DB 184 HVERELYPKEMGKRVVTSVGTQEYIDPPIRISNKGSRMGLEALEFWRGADVHTV 243
QY 241 TKGSKVAFIRIKUKVETVEMLSAIENELRSKKYDVVMAAASVDFRPKIKAGKIKS 300
DB 244 SKPSGMSLPNLYK-EIRVMSVEDMMKAVLYEI-GKGCDFVSSAAAADFTVDAEK-KIKT 300
QY 301 GRSITIELVPXNPKIIDRIKTEIQPNVFLVGPKAET--SKEKLIIEGKQIERAKADLVVG 358
DB 301 APELVULK--KESPKIIEKVRKLYSG-HIIGFKAETGMSDELKVKASEKMDNLMWVA 358
QY 359 NTL--EAFGSEENQVVLIGRDTFKLPKMK-----KRELAERI 394
```

```
Db 359 NDVLERGMGTEDTRVLIL-----TPKQEWVEGLKQHVABRI 395
RESULT 7
Q9HH70      PRELIMINARY;          PRT;    382 AA.
AC Q9HH70;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE MTW1216.
GN Name=mtw1216;
OS Methanobacterium wolfei.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanobacter.
OX NCBI_TaxID=145261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429258; PubMed=11544247;
RX DOI=10.1128/JB.183.19.5788-5792.2001;
RA Luo Y., Pfister P., Leisinger T., Wasserfallen A.;
RT "The Genome of Archaeal Prophage paim100 Encodes the Lytic Enzyme
RT Responsible for Autolysis of Methanothermobacter wolfeii.";
RL J. Bacteriol. 183:5788-5792(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Luo Y.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF301375; AAG39941.1; -.
DR HSSP; Q9HAB8; IP90.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
SQ SEQUENCE 382 AA; 41207 MW;  8E792042DFDFA3AB CRC64;

Query Match      36.6%; Score 731; DB 2; Length 382;
Best Local Similarity 45.5%; Pred. No. 1e-33;
Matches 178; Conservative 65; Mismatches 126; Indels 22; Gaps 13;
QY 20 KIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAAATKIHPYANLPTGNPVITEITG 78
DB 2 RVLVCVTGSVAAEAVKLAELRRKGAEVKCFMSDADRIIHPYAMEFATGSPVLELTG 61
QY 79 PIEHVELAGHENKADLLVCPATANTISKIACGIDDTPTVTVTTPPH-IPIMIAPM 137
DB 62 EIEHVKY-----SDADLLVAPATANIIGKLAYKLADNPISLLLTAAAGRGTAIIMVPSM 116
QY 138 HETMYRHPVIRENIRLKLGVFIPRIEGRKAVASIDEIVYVIRIKLHKHTLLEGRKV 197
DB 117 HEAMYE--AAENIRKLKEGVAFLEPRMDEGKAKFPDIDTITLEALQASEGKMRGLV 174
QY 198 LVTAGATREYIDPPIRITNASSGKMGVALAEADPRGA-VTLIRTKGSVK-AFRIKIKL 255
DB 175 LVSLGTLTLEIDPVRVITNRSSGRMGLAVAREAVIQGADVTLVAGTVSDIPSQRTV-- 232
QY 256 KVETVEMLSAIENELRSKKYDVVMAAASVDFRPKIKAGKIKSGRSITIELVPXNPKI 315
DB 233 RAETAHAEAAVAELI--GEHDVFVSAASVDFRP-VYSEEEKISDSEITLRLKP-NPKI 288
QY 316 IDRIEIQPNVFLVGVFKAE--TSKEKLIIEGKQIERAKADLVVGN--TLEAFGSEENQV 371
DB 289 IRMARETHPEAFIVGFKAHGVSEELAAARKQIEDSVADVADVANDVSVSEGGSENNRA 348
QY 372 VLIGRDTFKLPKMKKRELAERIWDIEIKL 402
DB 349 IIVSGVT-ELPTMKKEELAGLIIGEIMKRL 378

RESULT 8
O27284
```

ID Q27284 PRELIMINARY; PRT; 386 AA.
 AC Q27284;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DE 01-JUN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Pantothenate metabolism flavoprotein.
 GN OrderedLocusNames=MT1216;
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155 (1997).
 DR EMBL; AS000889; AAB85705.1; -;
 DR PIR; D69029; D69029.
 DR InterPro; IPR005252; Cons hypoth521.
 DR InterPro; IPR007085; Dfp_C.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF04127; DFP; 1.
 DR Pfam; PF02441; Flavoprotein; 1.
 DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
 KW Complete proteome.
 SQ SEQUENCE 386 AA; 42010 MW; 56386EAB76244A75 CRC64;

 Query Match 36.6%; Score 731; DB 2; Length 386;
 Best Local Similarity 44.9%; Pred. No. 1e-33;
 Matches 176; Conservative 62; Mismatches 122; Indels 32; Gaps 13;

 QY 20 KIVXXPGSIALD-VKACEGLIRHGAHVHVMSEATKIIHPYANLPTGNPVITEITG 78
 DB 2 EILCVTGSVAIEAVKLARELGRQAEVKCFMSEDACRIIHPYAMFATGSKPVLELTG 61

 QY 79 FIEHVELAGEHENKADLILVCPATANTISKIAGIDDTPTVTVTVTAP-PIHIPIMIAPAM 137
 DB 62 EIEHVKYAG-ADLILVAPATANIIGKLYRLADNPISLSSLLTASGNGTIVVWPSM 116

 QY 138 HETMYRHPIVRENIERLKKLVGFIPIRIEGRKVASDEIVYRVKKLHKTKLEGRV 197
 DB 117 HEAMY-AAAAAENIMLKEGVIFIEPRMDGKAKFPDIDTIVLEAMRQTSSQRLQGRV 174

 QY 198 LVTAGATREVIDPIRITWASSGKMGVALAEADFRGA-VTLIRTKGSV-----KAFRI 250
 DB 175 LVSLGTYEPIDPVGRITRSSGKMGVALARRAYIEGADVTVVAGTVSVIEPQLRSFR- 233

 QY 251 RKIKLVETVEEMLSAENELRSKKYDVWIMAAVSDFRPKIAEKGKISGRSITIELVP 310
 DB 234 -----AETAENAEVRVELVAD--HDVFIISAAVADFKP-VYTERKISSSEBSFVELAP 284

 QY 311 XNPKIIRIKIOPNVFLVGFKAET--TSKEKLEEGKQRIERAKDLVGN--TLAPGS 366
 DB 285 -NPKVIGIAREINPEAFIVGFAEYVDNEALVESARKQIRESGVDWVANDSVVEGFGS 343

 QY 367 EENOVVLIGRDTFKELPKMKKRELAERIWI 398
 DB 344 DRNRALIVS-DMVTELPMEKEBELASIIIDEV 374

 RESULT 9
 Q980S0
 ID Q980S0 PRELIMINARY; PRT; 413 AA.
 AC Q980S0;

DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DE DNA/pantothenate metabolism flavoprotein (Dfp).
 GN Name=dpf; OrderedLocusNames=SSO0210;
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726; DOI=10.1073/pnas.141222098;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840 (2001).
 DR EMBL; AS006658; AAK40553.1; -;
 DR PIR; B90162; B90162.
 DR InterPro; IPR005252; Cons hypoth521.
 DR InterPro; IPR007085; Dfp_C.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF04127; DFP; 1.
 DR Pfam; PF02441; Flavoprotein; 1.
 DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
 KW Complete proteome.
 SQ SEQUENCE 413 AA; 45864 MW; 79569D38D49A5219 CRC64;

 Query Match 32.5%; Score 648.5; DB 2; Length 413;
 Best Local Similarity 39.7%; Pred. No. 5.3e-29;
 Matches 166; Conservative 81; Mismatches 144; Indels 27; Gaps 11;

 QY 1 MLHVKLIIVATSKRKLGVKXVXXPGSIA---ALDVKACEGLIRHGAHVHVMSEATK 57
 DB 5 MTHPSKKIIGSISNELADKKILLAVTGSVAIVKSLDL--ARSLMRNGAEVSVIISDAAK 62

 QY 58 ITHPYANLPTGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIAGIDDTPT 117
 DB 63 LISPEMFKWATGNVNTVTKLTGLDLEHVSLEADN---DVMIVAPSTANTVVKIAYGADPT 118

 QY 118 VTTVVTAPPHI---PIMAPAMHETMYRHPIVRENIERLKKLVGFIPIRIEGRKAV 173
 DB 119 IT---ATALNFVGMKKPLIIVFSMHLMQVISPQVADAADRLKRGVIEPIEIVGLAHY 175

 QY 174 ASIDEIVYRVIK-KLHKKTLEGRKVLVGTAGATREYIDPIRITWASSGKMGVALAEADP 232
 DB 176 PKLEYLTSRITSVLRGKDLSGFNILATAGPTREYLDVSFRFISNPSSGTMGISIANEAYF 235

 QY 233 RGA-VTLITKGSVKAFRIRKIKLVETVEEMLSAENELRSKKYDVWIMAAVSDFRPK 291
 DB 236 RGAKRVVICPISSEKTEPYKQIVVETTEEMLNEVVKSIENGKFNVLGAPADYKFK 295

 QY 292 IKAEGKIKSGRSITIELVPKNPKIIDRIKEIQNVFLVGFKAET--SKEKLEEGKQRIE 349
 DB 296 NKSDTKIDSHTEIPKVELETPKISEYIRKY--NVLLVGFSAETVNSDELEIKAKIMR 353

 QY 350 RAKADLVGNLTLE---AFGSEENQVVLIGRD-FTKELPKMKKRELAERIWI 402
 DB 354 RHGFDLIVANNVRRKIDIGFSSEYNEVIVIDKNGDVRKIPKNFKVTIARKILDIVKEQL 411

 RESULT 10
 Q976C4
 ID Q976C4 PRELIMINARY; PRT; 412 AA.
 AC Q976C4;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

```
DE Hypothetical protein ST0257.
GN OrderedLocusNames=ST0257;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anai A., Kosugi H., Hosoyana A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kuehida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagis M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Ref. 8:123-140(2001).
DR EMBL; AP000982; BAB5223.1; --
DR InterPro; IPR005252; Cons_hypoth51.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coABC dfp; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 412 AA; 45517 MW; E2B28B9F11338947 CRC64;

Query Match 31.9%; Score 636.5; DB 2; Length 412;
Best Local Similarity 40.6%; Pred. No. 2.5e-28;
Matches 168; Conservative 76; Mismatches 145; Indels 25; Gaps 12;

QY 1 MLHVKLIYATSKRLVGKKIVXXPGSIAAL-DVKACEGLRHGAEVHVMSEATKII 59
DB 2 VHPSSKIIIGETISSELLNKKILVGTVGSIVYKTIIDLVRSLMRLGAEVQVIMSKDASKIV 61

QY 60 HPIYANLPNGNVIETITGFIHVELAGHENKADLLVCPATANTISKIAGIDTPTVT 119
DB 62 SPTMFEWATNGPVITKISGIEHVTLAEDY----DAYVAPATANTISKVINGIADTPTIT 117

QY 120 TVVTTAFPHI----PTMIAPAMHETVYRHPVIRENLERIKLGIVFIFGRIBEGRAKVAS 175
DB 118 ----VTALNFIGLKKLILVPANHLPMYPTPQMINNKLNEIGVIVVEPFLVRDVAHPD 174

QY 176 IDEIV-YRVIKLHKHTLGKRLVLTAGATRYIDIPITWASSGKMGVALAEADFRG 234
DB 175 IEVLSNYISTLLRGKDLGLKIVVTAGPTREHLPVRFISNPSSGTMGVAIANEAYFRG 234

QY 235 A-VTLIRTKGSVKAIRIKLKVETVEMLSAIENELRSKKYDVVIAAASVDFRPKIT 293
DB 235 ADVLVHGLSTHPIKPYVKKVQVETDDEMANEVKKFVET-GYNIVILAGAPADYKFN 293

QY 294 ASGKTKS-GRSTIELVPKNPKIIDRIKEIQPNVFLVGFKAETSKS--KLIEGKRQIER 350
DB 294 ASKIDSHTETTVEL-EKTVKISSIRKY--NVFLVGFAAETVKDDNELIEKGKIKER 350

QY 351 AKADLVVGNLTLE----AFGSEENQVVLIGRDFTELPRKMKRELARIWDETEK 400
DB 351 HGFDLLIANNAASKDKIAFSDFNEVIVIGNNFIKKINKYKTKVIAREILDIVKQ 404

RESULT 11
DFF METJA
ID - DFF METJA STANDARD; PRT; 403 AA.
AC Q58323;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Coenzyme A biosynthesis bifunctional protein coABC (DNA/pantothenate
DE metabolism flavoprotein) [includes: Phosphopantothenoylecysteine
DE decarboxylase (EC 4.1.1.36) (PPDC) (CoAC); Phosphopantothenate--
DE cysteine ligase (EC 6.3.2.5) (Phosphopantothenoylecysteine synthase)

DE (PPC synthetase) (PPCS) (CoAB)].
GN Name=coABC; OrderedLocusNames=MJ0913;
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.P., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Catalyzes two steps in the biosynthesis of coenzyme A.
CC In the first step cysteine is conjugated to 4'-phosphopantothenate
CC to form 4'-phosphopantothenoylecysteine, in the latter compound is
CC decarboxylated to form 4'-phosphopantothene (By similarity).
CC -!- CATALYTIC ACTIVITY: N-[(R)-4'-phosphopantothenoylecysteine =
CC pantothene 4'-phosphate + CO(2).
CC -!- CATALYTIC ACTIVITY: CTP + (R)-4'-phosphopantothenoylecysteine
CC = CMP + PPI + N-[(R)-4'-phosphopantothenoylecysteine.
CC -!- COFACTOR: Binds 1 FMN per subunit (By similarity).
CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; second step.
CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; third step.
CC -!- SUBUNIT: Homododecamer; the coAB domains form homodimers.
CC -!- SIMILARITY: In the N-terminal section; belongs to the HPCD (homo-
CC oligomeric flavin containing Cys decarboxylase) superfamily.
CC -!- SIMILARITY: In the C-terminal section; belongs to the PPC
CC synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U67535; AAB98918.1; -.
PIR; A64414; A64414.
HSSP; P30197; 1G63.
TIGR; MJ0913; -.
InterPro; IPR011256; Bac_reg_effector.
InterPro; IPR005252; Cons_hypoth51.
InterPro; IPR007085; Dfp_C.
InterPro; IPR003382; Flavoprotein.
Pfam; PF04127; DFP; 1.
Pfam; PF02441; Flavoprotein; 1.
TIGRFAMs; TIGR00521; coABC dfp; 1.
KW Complete proteome; Flavoprotein; FMN; Hypothetical protein; Ligase;
KW Lyase.
FT CHAIN 1 195 Phosphopantothenoylecysteine
FT decarboxylase.
FT CHAIN 196 403 Phosphopantothenate--cysteine ligase.
FT ACT_SITE 87 87 By similarity.
FT SEQUENCE 403 AA; 45670 MW; B8B51AA81A115B94 CRC64;

Query Match 31.4%; Score 627.5; DB 1; Length 403;
Best Local Similarity 39.6%; Pred. No. 8e-28;
Matches 163; Conservative 82; Mismatches 138; Indels 29; Gaps 16;

QY 1 MLHVKLIYATSKRLVGKKIVXXPGSIAALDV-KACEGLRHGAEVHVMSEATKII 59
DB 5 IMHPTKLLKGTKSLLENKKILVAVTSSIAAETPKLMRELIRHGAEVYCIITEYTKII 64
```



```

QY 126 RPHIPIAMHETMYRHPVIRENIERLKKLVGFIPGRIEGRKAVSI-----DE 178
DB 113 YGRCIFVAPAMTMYKPIVQNNIKKLKDYGYIFEP--EKGRACDVGEGKLADTE 170
QY 179 IVYRVIKKL--HKKTLEGKRVLTAGATREYIDPIRFITNASSGKGVALAEADFRGA- 235
DB 171 LIYENIKSLYNNKRDLAGKVMVTAGPTIAPDPVRFITNHSKGVAIAEADRGAE 230
QY 236 VTLIITKGSVK-APRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKA 294
DB 231 VTLSGETSLKPPFGVDFI--KVSTNSEMMKVEKFGSG--DIVIKSAADVAKAKNYS 286
QY 295 EGGI-KSGRSITIELVPXNPKIIDRIKEIQNVFLVGFKAETSKEKLIBEGRKQIERAKA 353
DB 287 ELUKIKASDNINIEFKOM-DILKKLGEIKKKHQLVGFPAESN--DLIENAYKLTIRKUL 343
QY 354 DLVVGNLT----EAFGSEENQVVLIGRDTFK-ELPKMKKRELAERIWDIEIKLS 403
DB 344 DIVIVANDILSKDAGFASDENRVVILGSDGSKLELDKMSKRKVAENLFDLLSKRS 398

RESULT 14
Q8XJLO PRELIMINARY; PRT; 395 AA.
AC Q8XJLO;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-JUN-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Probable pantothenate metabolism flavoprotein.
GN OrderedLocNames=CP81746;
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13;
RX MEDLINE=21664373; PubMed=11792842; DOI=10.1073/pnas.022493799;
RA Shimizu T., Ohtani K., Hirakawa H., Onshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kunihara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003191; BAB81452.1; -.
DR HSSP; QSW55; IE20.
DR Pfam; PF04127; DFP; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 395 AA; 43399 MW; 7F66FE05FB09D153 CRC64;

Query Match 28.9%; Score 577; DB 2; Length 395;
Best Local Similarity 38.6%; Pred. No. 5.7e-25;
Matches 156; Conservative 81; Mismatches 129; Indels 36; Gaps 17;

QY 19 KIVVXXPGSIA---ALDYKACEGLIRHGAEVHVMSEATKIIHPYAWNLPTGNPVITE 75
DB 5 KCVVGVSGGVAVYKALDY--ISRLRKDDVEVHVIMTKSATBFTVPLSFQSLSQNNVITD 62
QY 76 I-----TGFIEHVELAGEHENKADLILVCPATANTISKIACGI-DTPTVTVVTTAFPHI 129
DB 63 MFAEPKAWELQHSILA---KKADMLIIVPATANIIGKVANGIADMLSTTIMAT---KA 115
QY 130 PIMIAPAMHETMYRHPVIRENIERLKKLVGFIPG---RI---EGRKAVASIDEIVRV 183
DB 116 PVVFCPAMTMYNENIVQNNISLKLGEYFIEPASGRGLACGDEGKGLQDTETIAEET 175
QY 184 IKLHK-KTLEGKRVLTAGATREYIDPIRFITNASSGKGVALAEADFRGA-VTLIRT 241
DB 176 LRLHSTKDLGKGVVTVAGTPIVDPVRIITNRSKGKGYIAEADRGAEVVLISG 235
QY 242 KGSVKAIRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI-KS 300
DB 115

```

```

DB 236 PTLRKNPGIKV-IDIKTNEDMFNAIKNEF--KDADIVIKSAADVAKAKNYSNEKIKKT 292
QY 301 GRSITIELVPXNPKIIDRIKEIQNVFLVGFKAETSKEKLIBEGRKQIERAKADLVVNGT 360
DB 293 GDDLNL-IFERDRDLKTLGDMKENQILVGFPAESSNLK--ENAKGLDRKNLDIVAND 349
QY 361 LE----AFGSEENQVVLIGRDTFK-ELPKMKKRELAERIWDIE 399
DB 350 ISKSETGPASDENKVITISKSGEEVSLKMSKREKAKNIFDIK 393

RESULT 15
Q67PR7 PRELIMINARY; PRT; 414 AA.
AC Q67PR7;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Phosphopantothenate-cysteine ligase.
GN ORFNames=STH1341;
OS Symbiobacterium thermophilum.
OC Bacteria; Actinobacteria; Symbiobacterium.
OX NCBI_TaxID=2734;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IAM14863;
RA Ueda K., Yamashita A., Ishikawa J., Shimada M., Watsuiji T.,
RA Morimura K., Ikeda H., Hattori M., Beppu T.;
RT "Complete genome sequence of an uncultured bacterium Symbiobacterium
RT thermophilum.";
RL Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP006840; BAD40326.1; -.
DR GO; GO:0016874; F:ligase activity; IEA.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Ligase.
SQ SEQUENCE 414 AA; 43953 MW; EE403E2A1C08C78C CRC64;

Query Match 28.6%; Score 571; DB 2; Length 414;
Best Local Similarity 37.2%; Pred. No. 1.3e-24;
Matches 153; Conservative 73; Mismatches 141; Indels 44; Gaps 15;

QY 16 LVGKIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEATKIIHPYAWNLPTGNPVIT 74
DB 2 LHGKTVLLGVSGGIAAYKAAEICSLVLKLGADVHVLMTEATRLVAPLTFQALSGNPKV 61
QY 75 -----EITGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDTPTVTVVTTAFPHI 129
DB 62 DLMGQVHGVHVDHVLVT---HKADLLIAPATANTIAIRLAGGHAGDWIT--VTALGVRC 115
QY 130 PIMIAPAMHETMYRHPVIRENIERLKKLVGFIPGRIEGR-----AKVASIDEIVY 181
DB 116 PVLVAPSETMETYANPLTQNLRLILAEGLWDIMEF--EEGRSLASGLMGKRLPEPAKIVE 173
QY 182 RVIKKLHK-----TLEGKRVLTAGATREYIDPIRFITNASSGKGVALAEAD 232
DB 174 RAVALLQAREGRRPFGGDLAGRRVLTAGTTREPDPVRFITNRSKGKGYIAEAD 233
QY 233 RGAVTLIRTKSVKAFIRIKIL-KVETVEMLSAIENELRSKKYDVVIMAAVSDFRPK 291
DB 234 RGA-EVLVSGTHLEPPAGVRLVRSVCQMLEACLAEPAA--DVAVAAPADYRPA 290
QY 292 IKAEGKI-KSGRSITIELVPXNPKIIDRI-KEIQNVFLVGFKAETSKEKLIBEGRKQIE 349
DB 291 TVAPSKIKKTGELITLV-KNPDIIAELGRKRKQVTVFAAET--DDLIANARKLA 347
QY 350 RAKADLVVGNLTLEA-----FGSEENQVVLIGRDTFKELPKMKKRELAERIWD 396
DB 348 DKNAVDFVANDVTAEGAGFGTETNRVFTVADAVDELPLLPKREVADRIID 398

```

Search completed: July 2, 2005, 12:24:38
Job time : 83.7442 secs

THIS PAGE TELI REMARK

THIS PAGE LEFT BLANK

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	527	26.4	405	14	US-10-156-761-14401		Sequence 14401, A
2	439	22.0	420	9	US-09-738-626-5274		Sequence 5274, Ap
3	439	22.0	422	17	US-10-494-675-98		Sequence 98, Appl
4	351	17.6	425	15	US-10-335-977-5521		Sequence 5521, Ap
5	260	13.0	183	16	US-10-474-776-675		Sequence 675, App
6	260	13.0	183	17	US-10-472-928-2474		Sequence 2474, Ap
7	247.5	12.4	304	15	US-10-335-977-5520		Sequence 5520, Ap
8	238	11.9	234	17	US-10-472-928-2472		Sequence 2472, Ap
9	219	11.0	154	18	US-10-617-320-4272		Sequence 4272, Ap
10	197	9.9	228	9	US-09-895-913A-66		Sequence 66, Appl
11	165.5	8.3	309	15	US-10-376-774-2001		Sequence 2001, Ap

QY 20 KIVXXXPGSIAALDVKACEGLIR---HGAEVHAVMSEAA TKIIHPYA WNLPTGNPVITEI 76
 |:| | | | | | | | | : | : | : | : | :
Db 10 KWLVGVSSGIAA--YKACELLRLRTESGH DVVWPTGSALHFVGAA TWSALS GNPVSTEV 67

Db 67 IREITNTKOLL--IEMQERVQDYQVLIHSMVSDYTPVYMTGLEEVQASSNLKEFLSKQN 124

Qy 294 AEGKIKSGRITELVPXNPKIIDRIKBIQNPVFLVGFK--AETSKEKLIBEGKQIERA 351

Db 125 HQAKISSTDEVQVLFLLKTKPKIISLVKEMWNPITHLIGFKLLVDVTDHLVDIARKSLIKN 184

Qy 352 KADLVVGNLTLEAFSGSEENVVLIGRDTFKELPKMKKRELAERIWDEIE 399

Db 185 QADLIITANDLTQISADQHRFAIFVEKNQLOTV--QTKEEIAELLELEKIQ 230

RESULT 9

US-10-617-320-4272

; Sequence 4272, Application US/10617320

; Publication No. US20050136404A1

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID

; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC THERAPEUTICS

; NUMBER OF SEQUENCES: 5206

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: <Unknown>

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: <Unknown>

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/617,320

; FILING DATE: 10-Jul-2003

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,433

; FILING DATE: 30-Jun-1998

; APPLICATION NUMBER: 60/085131

; FILING DATE: May 12, 1998

; APPLICATION NUMBER: 60/051553

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-011

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 4272:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 154 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: Streptococcus pneumoniae

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (B) LOCATION 1...154

; SEQUENCE DESCRIPTION: SEQ ID NO: 4272:

US-10-617-320-4272

Query Match 11.0%; Score 219; DB 18; Length 154;

Best Local Similarity 36.9%; Pred. No. 9.6e-11;

Matches 52; Conservative 25; Mismatches 54; Indels 10; Gaps 3;

Qy 21 IVXXXPGSIAAL-DVKACEGLIRHGAHVHVMSEAATKIHPYAMNLTGNPVITEI--- 76

Db 16 ILLAVTGSIASYKSAIDLVSLLKKQGHQVTLMTQATFEIQPLTLQVLSQNPVHLDVMKE 75

Qy 77 --TGFIEHVELAGEHENKADLILVCPATANTISKIACIGDDTPVTIVVTTFAPFHIPIMIA 134

Db 76 PYPDQVNHIELG---KKADLFLIVVPATANTITAKLAHGFADNMVSTALALPSPHPIKLI 131

Qy 135 PAMHETMYRHPVIVRENIERLK 155

Db 132 PAMNTQYDHPVTQNNLKTLE 152

RESULT 10

US-09-895-913A-66

; Sequence 66, Application US/09895913A

; Patent No. US20020160456A1

; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold

; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles

; APPLICANT: Tomb, Jean Francois

; APPLICANT: Oomen, Raymond P.

; TITLE OF INVENTION: Identification of Polynucleotides

; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the

; FILE REFERENCE: 06132/043002

; CURRENT APPLICATION NUMBER: US/09/895,913A

; CURRENT FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 08/881,227

; PRIOR FILING DATE: 1997-06-24

; NUMBER OF SEQ ID NOS: 368

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 66

; LENGTH: 228

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-09-895-913A-66

Query Match 9.9%; Score 197; DB 9; Length 228;

Best Local Similarity 28.3%; Pred. No. 1.3e-08;

Matches 63; Conservative 44; Mismatches 88; Indels 28; Gaps 7;

Qy 14 RKLVGKKIVXXPGSIAAL-DVKACEGLIRHGAHVHVMSEAATKIHPYAMNLTGNPV 72

Db 12 RLEENKRVLLVSGSIAAYKSLVLLFPKSGASIQVVMKGAKKFKPLSPALSHHKV 71

Qy 73 ITEITGFIHVELAGEHEN-----KADLILVCPATANTISKIACIGDDTPVT-T 120

Db 72 LHD-----RNEKYYVNHQNALHNIACAANADLLIFAPLSTNSLSKIAHALADNIVSAT 126

Qy 121 VTTAFPHIPIMTAPAMHETMYRHPVIVRENIERLKKGVEFFIGPR-----IEEGRAKVA 174

Db 127 FLACASPKI---LAPSMNTNMLNSPITQSNLKRKLDNSHILDTKNALLACDTKGGGAMA 183

Qy 175 SIDEIVYRVIKKLHKKT-LEGKRVLVTAGATREYIDPIRFITN 216

Db 184 EPLEILFKAAQTLTKDAYENREVIVMGASIEKIDSVRTISN 226

RESULT 11

US-10-276-774-2001

; Sequence 2001, Application US/10276774

; Publication No. US20040053245A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; APPLICANT: Tang, Y, Tom et al

; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides

; FILE REFERENCE: 21272-030

; CURRENT APPLICATION NUMBER: US/10/276,774

; CURRENT FILING DATE: 2002-11-18

; PRIOR APPLICATION NUMBER: 09/560,875

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 09/496,914

; PRIOR FILING DATE: 2000-02-03

; NUMBER OF SEQ ID NOS: 2700

; SOFTWARE: Custom

; SEQ ID NO 2001

; LENGTH: 309

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (1)..(309)

; OTHER INFORMATION: Xaa = any amino acid or nothing

US-10-276-774-2001

Query Match 8.3%; Score 165.5; DB 15; Length 309;

Best Local Similarity 28.1%; Pred. No. 1.1e-05;

Matches 50; Conservative 30; Mismatches 65; Indels 33; Gaps 7;

QY 27 GSIALDVKACGLRTH--GAEVHVMSEAATKIHPY-----ANKL--PTGNP 71

DB 61 GSVAALKPLLVSKLLDIPGLEVAVVVTERAKHFYSPQDIPVTLVSDADDEWEMWKS RSDP 120

QY 72 VITEITGFTHEVELAGEHENKADLILVCPATANTISKIACGIDDPVTVVTTTAPPHIPI 131

DB 121 VL-----HIDL-----RRWADLLVAPLDANTLGKVASGICDNLITCVMAWDRSKPL 168

QY 132 MIAPAMHETMYRHPVIVRENIERLKKLG---VFIFGRI---BEGRAKVASIDEIVYRV 183

DB 169 LFCPAMNTAMWEHPITAQQVDQLKAFGYVEIPCVAKKLVCGDEGLGAMAEEVGTIVDKV 226

RESULT 12

US-10-871-385-7

; Sequence 7, Application US/10871385

; Publication No. US2005007957A1

; GENERAL INFORMATION:

; APPLICANT: INCYTE PHARMACEUTICALS, INC.

; APPLICANT: HILLMAN, Jennifer L.

; APPLICANT: BANDMAN, Olga

; APPLICANT: TANG, Y. Tom

; APPLICANT: YUE, Henry

; APPLICANT: LAL, Preeti

; APPLICANT: CORLEY, Neil C.

; APPLICANT: GUEGLER, Karl J.

; APPLICANT: PATTERSON, Chandra

; TITLE OF INVENTION: HUMAN RECEPTOR MOLECULES

; FILE REFERENCE: PF-0516 PCT

; CURRENT APPLICATION NUMBER: US/10/871,385

; CURRENT FILING DATE: 2004-06-21

; PRIOR APPLICATION NUMBER: 09/071,822

; PRIOR FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: Perl Program

; SEQ ID NO 7

; LENGTH: 326

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; OTHER INFORMATION: Incyte clone 2022379

US-10-871-385-7

Query Match

Best Local Similarity 28.1%; Pred. No. 1.2e-05;

Matches 50; Conservative 30; Mismatches 65; Indels 33; Gaps 7;

QY 27 GSIALDVKACGLRTH--GAEVHVMSEAATKIHPY-----ANKL--PTGNP 71

DB 26 GSVAALKPLLVSKLLDIPGLEVAVVVTERAKHFYSPQDIPVTLVSDADDEWEMWKS RSDP 85

QY 72 VITEITGFTHEVELAGEHENKADLILVCPATANTISKIACGIDDPVTVVTTTAPPHIPI 131

DB 86 VL-----HIDL-----RRWADLLVAPLDANTLGKVASGICDNLITCVMAWDRSKPL 133

QY 132 MIAPAMHETMYRHPVIVRENIERLKKLG---VFIFGRI---BEGRAKVASIDEIVYRV 183

DB 134 LFCPAMNTAMWEHPITAQQVDQLKAFGYVEIPCVAKKLVCGDEGLGAMAEEVGTIVDKV 191

RESULT 13

US-10-437-963-133863

; Sequence 133863, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 133863

; LENGTH: 220

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_35693C.1.pap

US-10-437-963-133863

Query Match

Best Local Similarity 27.3%; Pred. No. 1.4e-05;

Matches 42; Conservative 36; Mismatches 61; Indels 15; Gaps 3;

QY 20 KIVXXPGSIALDVKACEGLIRHCAEVHVMSEAATKIHPYANLPTGNP VITE--- 75

DB 22 RVLLAASGSVAATKPFSLCRSFSEWAEVRAVATKASLHPID--RTSLPSNIIYTDDBW 79

QY 76 -----ITGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVVTTTAPPHIP 130

DB 80 STWKIGDEVHLIEL-----RKWADIMVTAPLSANTLAKIAGGLCDNLLTCIVRAWDYSKP 135

QY 131 MIAPAMHETMYRHPVIVRENIERLKKLGVEFVIGP 164

DB 136 LfVAPAMNTFMWNPFTSRHLETINLLGISLVPP 169

RESULT 14

US-10-425-115-322884

; Sequence 322884, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 322884

; LENGTH: 220

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577_57533C.1.pap

US-10-425-115-322884

Query Match

Best Local Similarity 28.3%; Pred. No. 1.4e-05;

Matches 47; Conservative 35; Mismatches 63; Indels 21; Gaps 5;

[illegible]

RESULT 15

```

US-10-425-115-322885
; Sequence 322885, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 322885
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_57534C.1.pep
US-10-425-115-322885

```

Query Match . 8.1%; Score 161.5; DB 16; Length 220;
Best Local Similarity 28.3%; Pred. No. 1.4e-05;
Matches 47; Conservative 35; Mismatches 63; Indels 21; Gaps 5;

Qy		9	YATKGRKLVGKKIVXXPGSIALDLVKAEGELIRHGAEVHAVMSEAA TKIIH- PYAWNLP	67
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	15	YSQPSR----	PRVLLAASGSAAIKF-----ESLCRSFGEWADVRAVATTSSLHFVDRSSLP	67
	:	:	:	:
	:	:	:	:
	:	:	:	:
Qy	68	TGNPVITE-----	ITGFIEHVFLAGEHNKADILLVCPATANTISKIAGIDDPV	118
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	68	SGIVLYTDDDEWSWKKGIDVELHIETL----	RKWADVVIAPLSANTSANTAKIAGLCDNLL	123
	:	:	:	:
	:	:	:	:
	:	:	:	:
Qy	119	TTVVTTAFPHIIMIPAMAHETWRHPPIRVENIERLKLVGEFTGP	164	
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	124	TCGVRAWDSYKSLFPVAPAMNTLMWNPFTESHLHTINQLGALIPP	169	
	:	:	:	:
	:	:	:	:
	:	:	:	:

Search completed: July 2, 2005, 12:34:04
Job time : 81.0233 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 11:06:33 ; Search time 85.0698 Seconds
(without alignments)
1832.195 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 1998
Sequence: 1 MLHHVKLIYATKSRKLVGKK.....KKMKRELAEIRINDEIEKXLS 403

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1985	99.3	403	2 AAW72844	Polymers
2	1649.5	82.6	401	4 AAB96653	Putative
3	1574.5	78.8	406	8 ADN46559	Thermococ
4	509	25.5	376	7 ADM26107	Hyperther
5	494	24.7	399	5 ABB47692	Listeria
6	489	24.5	405	7 ABO75311	Pseudomon
7	466	23.3	418	7 ADF07150	Bacterial
8	461.5	23.1	399	4 AAG82741	S. epider
9	461.5	23.1	400	5 ABP39940	Staphyloc
10	461.5	23.1	400	8 ADS06229	Staphyloc
11	458.5	22.9	431	8 ADL05850	M. catarr
12	457	22.9	394	6 ABP80942	N. gonorr
13	453.5	22.7	401	6 ABM71249	Staphyloc
14	444	22.2	401	6 ADB10278	Alloioococ
15	444	22.2	401	8 ADJ27133	Alloioococ
16	444	22.2	422	4 AAB79946	Corynebac
17	439	22.0	420	4 AAG91520	C. glutami
18	439	22.0	422	7 ADD13531	C. glutami
19	439	22.0	427	6 ADA33918	Acinetoba
20	424.5	21.2	430	6 ABR63491	E coli fl
21	424.5	21.2	430	6 ABR63492	E coli fl
22	422	21.1	414	6 ABM65535	Propionib
23	418.5	20.9	398	7 ABO64139	Klebsiell
24	410	20.5	428	6 ABM68817	Photornab
25	264	13.2	184	8 ADS51776	Pantothene

26	262	13.1	183	8 ADK48063	Adk48063 Streptoco
27	260	13.0	183	6 ABU01661	Abu01661 S. pneumo
28	260	13.0	183	6 ABP81597	Abp81597 Streptoco
29	260	13.0	184	8 ADS51774	AdS51774 Pantothene
30	254.5	12.7	299	5 ABU51464	Abu51464 Helicobac
31	244	12.2	180	5 ABP28493	Abp28493 Streptoco
32	244	12.2	230	5 ABP30330	Abp30330 Streptoco
33	244	12.2	233	5 ABP28495	Abp28495 Streptoco
34	240	12.0	304	4 AAU67705	Aau67705 Propionib
35	240	12.0	304	6 ABM64224	Abm64224 Propionib
36	238	11.9	234	5 ABU01660	Abu01660 S. pneumo
37	237	11.9	230	5 ABP28496	Abp28496 Streptoco
38	236	11.8	252	8 ADK48064	Adk48064 Streptoco
39	235.5	11.8	245	7 ADC94821	Adc94821 E. faeciu
40	235	11.8	188	7 ADC97191	Adc97191 E. faeciu
41	233	11.7	262	7 ADH86020	Adh86020 Enterococ
42	232	11.6	181	5 ABP28494	Abp28494 Streptoco
43	224.5	11.2	231	5 ABB53883	Abb53883 Lactococc
44	223	11.2	178	5 ABB53882	Abb53882 Lactococc
45	219	11.0	154	8 ADR95637	Adr95637 Novel S.

ALIGNMENTS

RESULT 1
AAW72844
ID AAW72844 standard; protein; 403 AA.
XX
AC AAW72844;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P50 component.
XX
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
KW sequencing; replication.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Peptide 1..31
FT Misc-difference 23 /note= "N-terminal peptide used to generate primers"
FT Misc-difference 24 /note= "encoded by NNN"
FT Misc-difference 25 /note= "encoded by NNN"
FT Peptide 275..291 /note= "encoded by NNN"
FT Misc-difference 310 /note= "Internal peptide used to generate primers"
FT Misc-difference 311 /note= "encoded by CCN"
FT Peptide 353..376 /note= "encoded by NNN"
FT Misc-difference 376 /note= "Internal peptide used to generate primers"
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
XX
(STRA-) STRATAGENE.
XX
PI Hogrefe H, Hansen CJ;
XX
DR WPI; 1998-542284/46.

DR N-PSDB; AAV63859.
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 XX
 PS Claim 46; Page 36; 161pp; English.
 XX
 CC This is the amino acid sequence of the P50 component of the polymerase
 CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. The sequence is
 CC predicted from a DNA sequence (see AAV63859) obtained from genomic DNA by
 CC PCR. P50 and P45 (see AAV72847) are the predominant components of PEF,
 CC which acts to enhance the activity of P. furiosus DNA polymerase, thereby
 CC providing replication products of greater length and purity. P50 is
 CC similar in structure to a bacterial flavoprotein. The invention provides
 CC novel extracts, proteins and complexes that improve the polymerisation
 CC activity of nucleic acid polymerases, as well as DNA constructs and
 CC antibodies raised against P45 or P50. Also included are methods for
 CC identifying compositions with polymerase enhancing activity, for
 CC purifying and using these compositions, and specific extracts, proteins
 CC and complexes that function to enhance polymerase activity. Nucleic acid
 CC polymerase reactions can be enhanced (claimed) by mixing a nucleic acid
 CC template, at least 1 polymerase and a composition having polymerase
 CC enhancing activity. Kits are provided for replicating nucleic acids. The
 CC kits can be used in site-directed mutagenesis, nucleic acid sequencing or
 CC amplification (preferably PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 XX Sequence 403 AA;

Query Match 99.3%; Score 1985; DB 2; Length 403;
 Best Local Similarity 99.8%; Pred. No. 4.3e-184; Indels 0; Gaps 0;
 Matches 402; Conservative 1; Mismatches 0;
 QY 1 MLHHVKLIYATKSRKLVGKKIVXXXPGSIAALDVKACEGLIRHGAHVHVMSEATKIIH 60
 DB 1 MLHHVKLIYATKSRKLVGKKIVXXXPGSIAALDVKACEGLIRHGAHVHVMSEATKIIH 60
 QY 61 PYAWNLPTGNPVITTEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVT 120
 DB 61 PYAWNLPTGNPVITTEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVT 120
 QY 121 VVTTFAPPHIPIMAPAMHETMYRHPIVRENIERLKLGVETFGPRIEGRKAVASIDEIV 180
 DB 121 VVTTFAPPHIPIMAPAMHETMYRHPIVRENIERLKLGVETFGPRIEGRKAVASIDEIV 180
 QY 181 YRVIKLHKHTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGAVTLIR 240
 DB 181 YRVIKLHKHTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGAVTLIR 240
 QY 241 TKGSVKAPRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEKGKIS 300
 DB 241 TKGSVKAPRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEKGKIS 300
 QY 301 GRSITIELVPXNPKIIDRIKETOPNVFLVGFKAETSKEKLIIEGKQIERAKADLVVGN 360
 DB 301 GRSITIELVPXNPKIIDRIKETOPNVFLVGFKAETSKEKLIIEGKQIERAKADLVVGN 360
 QY 361 LEAFGSEENQVLIIGRDFTKELPKMKKRELAERIWDEIEKXLS 403
 DB 361 LEAFGSEENQVLIIGRDFTKELPKMKKRELAERIWDEIEKXLS 403

RESULT 2
 AAB96653
 ID AAB96653 standard; protein; 401 AA.
 XX
 AC AAB96653;
 XX
 DT 29-OCT-2001 (first entry)
 XX Putative P. abyssi flavoprotein involved in panthothenate metabolism.
 DE
 XX

KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX *Pyrococcus abyssi*.

XX FR2792651-Al.

XX 27-OCT-2000.

XX 21-APR-1999; 99FR-00005034.

XX 21-APR-1999; 99FR-00005034.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX (IFRE-) IFREMER INST FR RECH EXPL MER.

XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

PI Querellou J, Weissenbach J, Saurin W, Heilig R;

XX WPI; 2001-126236/14.

XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins useful in industry.

PS Claim 7; Page 1403-1404; 1657pp; French.

XX The present invention relates to the genomic sequence of *Pyrococcus abyssi* (see AAF8431 and AAH41223-7) and *P. abyssi* proteins. *P. abyssi* is a hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal vents. The present invention is one such *P. abyssi* protein. The proteins of the present invention have various potential industrial uses, since the proteins are stable at very high temperatures, some up to 110 degrees centigrade. Note: This patent is in the same patent family as WO200005052, which contains additional sequences as shown in AAB99132-AAB99143, AAH75903-AAH75920 and AAG68436

XX Sequence 401 AA;

Query Match 82.6%; Score 1649.5; DB 4; Length 401;

Best Local Similarity 82.9%; Pred. No. 2e-151;

Matches 335; Conservative 31; Mismatches 33; Indels 5; Gaps 4;

QY 1 MLHHVKLIYATKSRKLVGKKIVXXXPGSIAALD-VKACEGLIRHGAHVHVMSEATKII 59

DB 1 MLSHKLIYATKSRKLVGKKIVLAIPGSIAAVECVKLARELIRHGAHVHVMSPSATKII 60

QY 60 HPYAWNLPTGNPVITTEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVT 119

DB 61 HPYAMEFATGNPVITTEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVT 120

QY 120 TTTTTFAPPHIPIMAPAMHETMYRHPIVRENIERLKLGVETFGPRIEGRKAVASIDEI 179

DB 121 TTTTTFAPPHIPIMAPAMHESMYKPIVRENIEKLVGVEFGPRIEGRKAVATIDEI 180

QY 180 YRVIKLHKHTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-VTL 238

DB 181 YRVIKLHKHTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-VTL 240

QY 239 IRTKGSVKAPRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEKGI 298

DB 241 IRTGRSVKSFVENQI--EVETVEEMLGAIEEELTKKCYDVVIMAAVSDFRPKIKAEKGI 298

QY 299 KSGRSITIELVPXNPKIIDRIKETOPNVFLVGFKAETSKEKLIIEGKQIERAKADLVV 358

DB 299 KSDKSITIELVP-NPKIIRKIKIISDVFLVGFKAETSKEKLIIEGKQIESAGSDLVIG 357

QY 359 NTLEAFGSEENQVLIIGRDFTKELPKMKKRELAERIWDEIEKXL 402

DB 358 NTLEAFGSEESKVIVGKDFVKELPKMKKRELAERIWDEIEKXI 401

RESULT 3

ADN46559

ID ADN46559 standard; protein; 406 AA.

XX AC ADN46559;
XX AC
XX
XX
XX 01-JUL-2004 (first entry)
XX DT
XX DE Thermococcus kodakaraensis KOD1 protein sequence SeqID437.
XX DE
XX DE gene disruption; gene targeting; marker gene; transformation;
KW homologous recombination; hyperthermostable archaeobacterium; KOD1;
KW gene structure; gene function; enzyme activity; medicine;
KW forensic science; food; drug inspection; molecular biology; immunology.
XX
XX Thermococcus kodakaraensis.
XX
XX WO2004022736-A1.
XX
XX 18-MAR-2004.
XX PF 29-AUG-2003; 2003WO-IB003597.
XX PR 30-AUG-2002; 2002JP-00319011.
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX Imanaka T, Atomi H;
XX WPI; 2004-257583/24.
XX
XX Method for disrupting targeted gene in genome of organism particularly
PT thermostable bacterium and with genome chips for analysis, applicable in
PT studying gene structure and functions.
XX
XX Claim 9; SEQ ID NO 437; 598pp; Japanese.
XX
XX This invention relates to a novel method for targeting disruption of an
CC arbitrary gene in a genome of an organism which comprises providing the
CC whole sequential data of the genome of such organism, selecting at least
CC 1 arbitrary region in the sequence, providing a vector that contains a
CC sequence homologous with the selected region and a marker gene,
CC transformation, and homologous recombination. The genome is preferably
CC the genome of a hyperthermostable archaeobacterium, particularly
CC Thermococcus kodakaraensis KOD1. The method is for targeting the
CC disruption of a gene in the genome of an organism, which is applicable in
CC studying gene structure and functions as well as enzyme activities of
CC encoded proteins and useful in medicine, forensic science, food or drug
CC inspection, molecular biology and immunology. With this method, the
CC disruption of a gene at an arbitrary position in a genome can be achieved
CC efficiently and reliably. The present sequence is that of a protein
CC encoded by the genome of Thermococcus kodakaraensis which was derived
CC using the method of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 406 AA;
SQ
Query Match 78.8%; Score 1574.5; DB 8; Length 406;
Best Local Similarity 78.8%; Pred. No. 4.1e-144;
Matches 320; Conservative 37; Mismatches 42; Indels 7; Gaps 5;
QY 1 MLHHVKLIYATSRKLVGKKIVXXXPSTAAALD-VKACEGLIRHGAHVHVMSEAAATKII 59
DB 4 MLHHVKLIYATSRKLVGKKIIVLAIPGSTAAVECVKRLARELIRHGAHVHVMSEAAATKII 63
QY 60 HPYANNLPTGNVITTEITGFIHVELAGHENKADILVCPATANTISKIACIDDTPTV 119
DB 64 HPYAMEFATGNPVVTEITGFIHVELAGHENKADILVCPATANTISKIACIDDTPTV 123
QY 120 TVVTTAFPIPIAMAPMHTMYRHPIVRENIERLKKLGVEFTGPRIEGRKAVASIDRI 179
DB 124 TVVTTAFPIPIAMAPMHTMYRHPIVRENIERLKKLGVEFTGPRIEGRKAVASIEI 183
QY 180 VYRVIKKLHKHTLEGKRVLVTAAGATREYIDPIRITNASSGKMGVALAEAEADFRGA-VTL 238

DB 184 VYRVIKKLHPKSLGKRVLVTAAGATREYIDPIRITNASSGKMGVALAEAEAEFRGAETVL 243
QY 239 IRTKGSVKAFIRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAAVSDFRPKIKAEKGI 298
DB 244 IRTKSSVPSFVENQI--EVEVTEEMLEAJESLKGKKYDVVVLAAAVSDFRVKNKADVKI 301
QY 299 KSGRITIELVPXPKIIDRIKEIQNPVFLVGFKAET--SKEKLTIEGKQIRAKADLV 356
DB 302 KSGQLVLVLELP-TPKIIDRVKELQPGVFLVGFKAETGLSEBELTISAARKQIRAGSDLV 360
QY 357 VGNLTLEAFSGSENOVVLIGRDTFKELPKMKKRELAERIWDEIEKXL 402
DB 361 VANTLKAFGSENEVVLVGRDPAKLPRTWKRELAERLWDEIEKML 406
RESULT 4
ADM26107
ID ADM26107 standard; protein; 376 AA.
XX
XX ADM26107;
XX
XX 20-MAY-2004 (first entry)
XX
XX Hyperthermophile Methanopyrus kandleri protein #713.
XX
XX hyperthermophile; protein stability enhancement;
KW protein activity enhancement.
XX
XX Methanopyrus kandleri.
XX
XX WO2003076575-A2.
XX
XX 18-SEP-2003.
XX
XX 04-MAR-2003; 2003WO-US006664.
XX
XX 04-MAR-2002; 2002US-0361742P.
PR 14-MAY-2002; 2002US-0380423P.
PR 16-SEP-2002; 2002US-0410974P.
XX (FIDE-) FIDELITY SYSTEMS INC.
PA (MALY/) MALYKH A.
XX
XX Slesarev AI, Pavlov A, Pavlova N, Kozyavkin S;
PI WPI; 2003-748383/70.
DR N-PSDB; ADM27081.
XX
XX New isolated nucleic acids encoding any of about 1700 Methanopyrus
PT kandleri proteins, and the encoded proteins, useful as a medicaments or
PT as diagnostic agents.
XX
XX Claim 31; SEQ ID NO 713; 1023pp; English.
XX
XX The invention comprises the amino acid sequence of proteins from the
CC hyperthermophile Methanopyrus kandleri, the invention also comprises the
CC complete genome from Methanopyrus kandleri. The Methanopyrus kandleri
CC proteins of the invention are useful for enhancing the stability and/or
CC activity of other proteins. The Methanopyrus kandleri genome is useful in
CC a variety of diagnostic and analytical methods. The present amino acid
CC sequence represents a Methanopyrus kandleri protein of the invention.
XX
XX Sequence 376 AA;
SQ
Query Match 25.5%; Score 509; DB 7; Length 376;
Best Local Similarity 37.9%; Pred. No. 2.1e-40;
Matches 149; Conservative 59; Mismatches 149; Indels 36; Gaps 15;
QY 21 IVXXXPSTAAAL--DVKACEGLIRHGAHVHVMSEAAATKIIHPYANNLPTGNPVITEIT 77
DB 4 VVVCVTSVAAYRAPDV--CRELVGRHVRVVAEEALRFVKDGLGF-AAEEVIFRLT 60

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
DR N-PSDB; ABD0882.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 24057; 455pp; English.
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences AB067826-
CC AB084396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 405 AA;
Query Match 24.5%; Score 489; DB 7; Length 405;
Best Local Similarity 34.8%; Pred. No. 2e-38;
Matches 136; Conservative 77; Mismatches 132; Indels 46; Gaps 16;
QY 14 RLVGKKIVXXPGSIAALDVKACEGLIR-----HGAEVHVMSEATKLIHPYAVNLPTG 69
DB 5 QRLYRKRIVLVGGGGIA--YKSAE-LVRLRDQGAERVVMVTOGGREIFITPLTQLASG 61
QY 70 NPVITEI-----TGFIEHVELAGEHENKADLIILVCPATANTISKIACGIDDTPTVTWVTT 124
DB 62 HPVHTDLLDPAAEAAMGHIELA---RWADLVLIAPATADLMARLVQGVANDLLTTLVLA 117
QY 125 APHPIMTAPAMHETMYRHPVIRENIERLKKLGVEFIGPRI-----BEGRAKVASIDE 178
DB 118 T--DAQIALAPAMNQAMWRDTATQANAELLRQRFHLFGPAAGSQACGDVGLGRMLEABE 175
QY 179 IYVRVVKLHKHKKLEGRVLTAGATREYIDPIRFTLNASSGKMGVALAEEDFRGA-VT 237
DB 176 LAQRAADCFQORALTGVHVLITAGTQENIDPVRYITNHSKGKMGFALAEAAVEAGARVT 235
QY 238 LIRTKGSV---KAFIRKIKLVETVEEMLSAIENELRSKKYDVWVMAAAVSDFRPKKA 294
DB 236 LV--TGPVHLPTDRVQRV--DVVSARDMLACEAEM---PCDLLTASAADYREVA 288
QY 295 EGKIK---SGRSITIELVPXPKIIDRIKIQPNVFLVGFKAETSKELIEGKQRIER 350
DB 289 AHNKKKDPSTSGELLQLV--RNPDIILATLAQREDRPFVSGFAAET--ENLLDYAARKLKD 345
QY 351 AKADLVVGNLTLE---AFGSEENQVNLGRD 377
DB 346 KNLDLIVANDVANPSIGFNSDENAITVIDRD 376
RESULT 7
ADP07150
ID ADP07150 standard; protein; 418 AA.
XX
AC ADP07150;
XX
DT 12-FEB-2004 (first entry)
XX
DE Bacterial polypeptide #3263.
XX

KW Proteus mirabilis infection; bacterial infection; antibacterial;
XX immunostimulant.
OS Proteus mirabilis.
XX
PN US6605709-B1.
XX 12-AUG-2003.
XX
PF 05-APR-2000; 2000US-00543681.
XX
PR 09-APR-1999; 99US-0128706P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL;
XX WPI; 2003-895291/82.
DR N-PSDB; ADP02978.
DR
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX
PS Disclosure; SEQ ID NO 7435; 870pp; English.
XX
CC The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis,
CC a method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX
SQ Sequence 418 AA;
Query Match 23.3%; Score 466; DB 7; Length 418;
Best Local Similarity 34.2%; Pred. No. 3.7e-36;
Matches 149; Conservative 76; Mismatches 145; Indels 66; Gaps 20;
QY 5 VKL---IYATK---SRKLVGKKIVXXPGSIAALDVKACEGLIRH---CAEVHVMSEA 54
DB 3 VKLWTFIYAIRHFIMTTLHDKNIIIGISGIAA--YKAPE-LVRLRDKGAIYRVVWTPA 59
QY 55 ATKIHYPYAWNLPTGNPVITEI-----TGFIEHVELAGEHENKADLIILVCPATANTISKI 109
DB 60 AHAFVTPLSIQAVSGFPFVADLLDPAAEAAMGHIELG---KWADLILLAPATADLIARL 115
QY 110 AGCIDTPTVTVTVAFFPHIPMIAPAMHETMYRHPVIRENIERLKKLGVEFIGP----- 164
DB 116 RMGMANDLLTTLCLAS--AAPTAIPAMNQMYRATITQENLTALAQRCGLIWGPGSGS 173
QY 165 ---RIEGR---AKVASIDEIYVRVVKLHKHKKLEGRVLTAGATREYIDPIRFTTNA 217
DB 174 ACDVGPGRMLDPLELVALAEQOFAL-----QHDFFGKKITITAGPTREALDPVRFISNH 228
QY 218 SSGKMGVALAEAEADFRGA-VTLIR-----TKGSVKAFIRKIKLVETVEEMLSAIENE 270
DB 229 SSGKMGFAIAQAAQAQGAETVLIAGPVTLPACVK-----RIDVESAEQVMQVMDI 281
QY 271 LRSKKYDVWVMAAAVSDFRPKKAESKI-KSGRSITIELVPXPKIIDRI-KIQPNVFL 328
DB 282 AQSQ--DIFTCAAVADYRAKQVAPKIKKQSGDEVITMI-KNPDIIVASVYGRKVEHPPV 338
QY 329 VGFKAETSKELIEGKQRIERAKADLVVGNLTLE---AFGSEENQVNLGRDFTKELPK 384
DB 339 VGFAAET--QNVVEYARRKREKQKQLDLCANDVSLDKDAGFNSDNLHLIHWANGETRLPH 396

```
QY 385 MKKRELAEIRWDEIEK 400
Db 397 SSKTQLSHRLDLEIAK 412

RESULT 8
ID AAG82741 standard; protein; 399 AA.
XX
AC AAG82741;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2576.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
KW endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US030782.
XX
PR 09-NOV-1999; 99US-0164258P.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
XX
DR N-PSDB; AAH53591.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
XX
PS Claim 18; Page 675; 2188pp; English.
XX
AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX
SQ Sequence 399 AA;
Query Match 23.1%; Score 461.5; DB 4; Length 399;
Best Local Similarity 34.3%; Pred. No. 9.5e-36;
Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;
QY 19 KIVVXXPGSIAALD-VKACEGLIRHGAHVHVMSEATKIHPYANWLPTGNPVIT--- 74
Db 2 KHLIAVTGGIAAYKAIDITSKLIQSGYDVRVYMLSDHAQEFVTPLAFQAIRSNRPVYTNF 61
QY 75 --EITGFIEHVELAGEHENKADLIILVCPATANTISKIACGI-DDTPVTVTVTATPPIHI 131
```

```
Db 62 KEENPERIQHVSIG---DWADAIIVADPATANTIAKLSVGIADDLITSTLLATTPK--- 114
QY 132 MIAPAMHETMYRHPVIRENIERLKKLGVERFICP-----RIEGRKAVASIDE 178
Db 115 FVPAMNMYNNPRTHNMKVLSDQGYFFIEPGSGYLACGVAVAKRMEE--PMQLSVIN 173
QY 179 IVYRVIKKHKKTLEGKRVLTAGATREYIDPIREITNASSGKMGVALAEEDFRGA-VT 237
Db 174 KFTQOKNVVKSSFGKALVTAGPTVEIDPVRYVSNRSSGKMGYATAEARLDRKGAIVT 233
QY 238 LIRTKGSVKAFRIRKIK-LKVETVEEMLSALENELRSKKYDVVIMAAVSDPRPKIKARG 296
Db 234 LI--SGPHTLSLPEGINVVKVESADDMFQAVTE--RFAKQDIVIKAAAASVDYTPMDILBH 289
QY 297 KIK---SRSITIELVPANPKIIDRIKEIQPNVLGVGFKAETSKEKLEEGRKQIERAKA 353
Db 290 KKKQEGSLSVQFK---RTKDILKYLGENKTHQYLVGFAAET--QNIEQYALDKLRKNA 344
QY 354 DLV---VGNLTLEAFGSENOVVLIGRDFTK-ELPKMKKRELAERIWDBIE 399
Db 345 DVIISNNVGDTSIGFSSDDNLTWHFKNEKNKNIKKKKSALAHQIIEILE 395

RESULT 9
ABP39940
ID ABP39940 standard; protein; 400 AA.
XX
AC ABP39940;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4785.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-00134001.
XX
PR 14-AUG-1997; 97US-0055779P.
XX
PR 08-NOV-1997; 97US-0064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
WPI; 2002-381255/41.
XX
DR N-PSDB; ABN92485.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.
XX
PS Disclosure; SEQ ID NO 4785; 267pp; English.
XX
ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life cycle
CC or inhibit S. epidermidis infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site
XX
SQ Sequence 400 AA;
```

Query Match 23.1%; Score 461.5; DB 5; Length 400;
 Best Local Similarity 34.3%; Pred. No. 9.6e-36; Mismatches 148; Indels 47; Gaps 16;
 Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;

QY 19 KKIIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAATKIIHPYAWNLPTGNPVIT--- 74
 DB 3 KXILLAVTGGIAAYKAIDLTSKLIQSGYDVRVMSDHAQEFVTPFLAQISRNPVYNTF 62

QY 75 --BITGFIEHVELAGHENKADILVCPATANTISKIACGI-DDTPVTVTVTTAPPHIPI 131
 DB 63 KEENPEIQHVSIGL---DWADAIIVAPATANTIAKLSVGIADDLTITSTLLATTPK--- 115

QY 132 MIAPAMHETMYRHPVIRENIERLKLGVFEIGP-----RIEGRKVASIDE 178
 DB 116 FVAPANNVNNPNRTKHNKVLSDQGYFIEPGSYLACGYVAKGMEE-PMQILSVIN 174

QY 179 IVRVIKLHKHTLEGKRVLTAGATREYIDPIRFTITNASSGKMGVALAEADFRGA-VT 237
 DB 175 KFTQKQNVVKSFSFGKRALVTAGPTVEVDVRYVNSRSGKMGYALAEALDKGAIVT 234

QY 238 LIKTKGSVKAFIRKIK-LKVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAE 296
 DB 235 LI--SGPHTLSLPEGINVVKVESADDMFQAVTE--RFAKQDIVIKAAVSDYTPMDILEH 290

QY 297 KIK---SGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEGKQIERAKA 353
 DB 291 KLKQEGGLSVQPK---RTKDILKYLGENKTHQYLVGFAAET--QNIQYALDKLKRKNA 345

QY 354 DLV---VGNTEAFGEENQVVLIGRDFTK-ELPKMKKRELAERIWEIE 399
 DB 346 DVIISNNVGDTSIGFSSDDNLTMHFPKNEKNVKKGKSALAHQIIEILE 396

RESULT 10

ADSO6229
 ID ADSO6229 standard; protein; 400 AA.
 XX ADSO6229;
 AC ADSO6229;
 DT 04-NOV-2004 (first entry)
 XX
 DE Staphylococcus epidermis polypeptide seqid 5524.
 KW antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
 KW recombinant expression vector; infection; computer readable medium;
 KW computer based system.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN US2004147734-A1.
 XX
 PD 29-JUL-2004.
 XX
 PF 01-DEC-2003; 2003US-00724972.
 XX
 PR 08-NOV-1997; 97US-0064964P.
 PR 13-AUG-1998; 98US-00134001.
 PR 29-NOV-1999; 99US-00450969.
 XX
 XX (DOUC/) DOUCETTE-STAMM L.
 PA (BUSH/) BUSH D.
 XX
 XX Doucette-Stamm L, Bush D;
 XX
 XX WPI: 2004-580138/56.
 DR N-PSDB; ADS02457.
 DR
 XX
 PT New isolated polypeptide and encoding nucleic acid derived from
 PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
 PT treating an S. epidermidis bacterial infection.
 XX
 XX Claim 17; SEQ ID NO 5524; 741pp; English.
 PS
 XX

CC The invention describes an isolated nucleic acid comprising a nucleotide
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
 CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
 CC given in the specification. Also described are: a recombinant expression
 CC vector; a cell comprising a recombinant expression vector of (1);
 CC producing an S. epidermidis polypeptide; an isolated nucleic acid
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
 CC vaccine composition for prevention or treatment of an S. epidermidis
 CC infection, comprising a nucleic acid cited above and a carrier; treating
 CC a subject for S. epidermidis infection; a recombinant or substantially
 CC pure preparation of an S. epidermidis polypeptide or its fragment; a
 CC vaccine composition for prevention or treatment of an S. epidermidis
 CC infection; detecting the presence of a Staphylococcus nucleic acid in a
 CC sample; a computer readable medium having recorded in it the nucleotide
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
 CC system for identifying fragments of the Staphylococcus genome of
 CC commercial importance; a computer based system for identifying fragments
 CC of the Staphylococcus plasmids of commercial importance; identifying
 CC commercially important nucleic acid fragments of the Staphylococcus
 CC genome and/or plasmids; and identifying an expression modulating fragment
 CC of the Staphylococcus genome and/or plasmids. The methods and
 CC compositions of the present invention are useful for the diagnosis,
 CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
 CC infection. This is the amino acid sequence of a S. epidermis protein of
 CC the invention.

XX
 SQ Sequence 400 AA;

Query Match 23.1%; Score 461.5; DB 8; Length 400;
 Best Local Similarity 34.3%; Pred. No. 9.6e-36;
 Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;

QY 19 KKIIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAATKIIHPYAWNLPTGNPVIT--- 74
 DB 3 KXILLAVTGGIAAYKAIDLTSKLIQSGYDVRVMSDHAQEFVTPFLAQISRNPVYNTF 62

QY 75 --BITGFIEHVELAGHENKADILVCPATANTISKIACGI-DDTPVTVTVTTAPPHIPI 131
 DB 63 KEENPEIQHVSIGL---DWADAIIVAPATANTIAKLSVGIADDLTITSTLLATTPK--- 115

QY 132 MIAPAMHETMYRHPVIRENIERLKLGVFEIGP-----RIEGRKVASIDE 178
 DB 116 FVAPANNVNNPNRTKHNKVLSDQGYFIEPGSYLACGYVAKGMEE-PMQILSVIN 174

QY 179 IVRVIKLHKHTLEGKRVLTAGATREYIDPIRFTITNASSGKMGVALAEADFRGA-VT 237
 DB 175 KFTQKQNVVKSFSFGKRALVTAGPTVEVDVRYVNSRSGKMGYALAEALDKGAIVT 234

QY 238 LIKTKGSVKAFIRKIK-LKVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAE 296
 DB 235 LI--SGPHTLSLPEGINVVKVESADDMFQAVTE--RFAKQDIVIKAAVSDYTPMDILEH 290

QY 297 KIK---SGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEGKQIERAKA 353
 DB 291 KLKQEGGLSVQPK---RTKDILKYLGENKTHQYLVGFAAET--QNIQYALDKLKRKNA 345

QY 354 DLV---VGNTEAFGEENQVVLIGRDFTK-ELPKMKKRELAERIWEIE 399
 DB 346 DVIISNNVGDTSIGFSSDDNLTMHFPKNEKNVKKGKSALAHQIIEILE 396

RESULT 11
 ADSO5850
 ID ADSO5850 standard; protein; 431 AA.
 XX ADSO5850;
 AC ADSO5850;
 DT 06-MAY-2004 (first entry)
 XX
 DE M. catarrhalis protein #1616.
 XX
 XX Moraxella catarrhalis; infection.


```
XX OS Moraxella catarrhalis.
XX PN US6673910-B1.
XX PD 06-JAN-2004.
XX PF 04-APR-2000; 2000US-00540236.
XX PR 08-APR-1999; 99US-0128416P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton GL;
XX WPI; 2004-178127/17.
XX DR N-PSDB; ADL03930.
XX PT New nucleic acid encoding a Moraxella catarrhalis polypeptide, useful for
XX PT preparing a composition for diagnosing, preventing or treating infection
XX PT caused by Moraxella catarrhalis.
XX PS Disclosure; SEQ ID NO 3536; 429pp; English.
XX CC The invention relates to an isolated nucleic acid encoding an Moraxella
XX CC catarrhalis polypeptide. The nucleic acid is useful for preparing a
XX CC composition for diagnosing, preventing or treating infection caused by
XX CC Moraxella catarrhalis. The present sequence represents the amino acid
XX CC sequence of a M. catarrhalis protein.
XX SQ Sequence 431 AA;
Query Match 22.9%; Score 458.5; DB 8; Length 431;
Best Local Similarity 33.0%; Pred. No. 2.1e-35;
Matches 140; Conservative 88; Mismatches 147; Indels 49; Gaps 19;
QY 5 VKLIVATSKRLVGK--KIVXXPGSIAAL-DVKACEGLIRHGAHVAMVSEAATKIHP 61
DB 11 IMNVFAKQKQVNTQPKILLAVTGIIAAYKSAMLARLLIKSGCQVVMWTAGACEFITP 70
QY 62 YAWNLPNGPVITEITG-----FIEHVELAGHENKADLILVCPATANTISKIACGIDDT 116
DB 71 LTFQALTQGVHTKLLDETAERGGMGHITLA---KWADLVVVAPASANTIAKLAGGFADN 126
QY 117 PVTVVTTAFPHPIAMIAAMHETMYRHPVIVRENIERLKKLVGEFIPRIE-----GR 170
DB 127 LVTTVCLAT--DAPILIAPANMQMWAIVQDNLKLRQFGYHMTPDGSGQACGDVGL 184
QY 171 AKVASIDEIVYRVI---KKLH-KKTLGKRVLTAGATREYIDPIRFITNASSGKMGVAL 226
DB 185 GRLEPEDLCQQLAFDCKRHPQLPSLLGKTVITAGATVEPIDVRFSLNHSHTGKGYAL 244
QY 227 ABEDAFRGA-VTLIRTKG-SVKA-FRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAA 283
DB 245 ANACYHAGADVILVSKHVSLLKTPFGVR--RLDVGTAQMLKVCLSV--CQKADIFIAA 300
QY 284 AVSDFRPKIKAGSKI---KSGRSITTELVPXNPKIIDRIKEIQPNVFLVGFKAET----- 335
DB 301 AVADFKAASVADHKIKTKNQQTMTLQLI-KNPDVLATICTHTYDLCVGFKAETQDANN 359
QY 336 -SKEKLEEGKQIERAKADLVVGNLTLEAFGSEENOVVLIGRD---FTKELPKMKKREL 390
DB 360 CAKSKL---AAKQOLDIAVNDVSDKTI-GFGSENNAMTVFFAEQYDQMPQLPKAHKSKI 415
QY 391 AERI 394
DB 416 AEQL 419
RESULT 12
ID ABP80942
XX ABP80942 standard; protein; 394 AA.
```

```
AC ABP80942;
XX 07-MAR-2003 (first entry)
XX N. gonorrhoeae amino acid sequence SEQ ID 8414.
XX DE Antibacterial; infection; vaccine; gene therapy.
XX KW Neisseria gonorrhoeae.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB002069.
XX PR 12-FEB-2001; 2001GB-00003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
XX DR N-PSDB; ABZ41912.
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX PT medicament for treating or preventing N. gonorrhoeae infection.
XX PS Disclosure; Page 801; 815pp; English.
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
XX CC antibodies that specifically bind to the proteins. The composition
XX CC comprising the protein, nucleic acid or antibody is useful for the
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
XX CC infection, this may be in the form of a vaccine or gene therapy.
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX CC molecules of the invention
XX SQ Sequence 394 AA;
Query Match 22.9%; Score 457; DB 6; Length 394;
Best Local Similarity 33.7%; Pred. No. 2.6e-35;
Matches 138; Conservative 72; Mismatches 161; Indels 38; Gaps 15;
QY 17 VGKXIVXXPGSIAALDVKACEGLIR---HCAEVHVMSEAATKIHPYANLPTGNPV 72
DB 1 MGRHILLGVGTGIAA--YKSCE-LVRLKKQGHSTVVMRSRATFVSPLTTQALSGNPV 57
QY 73 ITEITG-----FIEHVELAGHENKADLILVCPATANTISKIACGIDDTFVTVTVTAPP 127
DB 58 LTDTHGGNGSGMEHINT---RNADVFLIAPASMTVAKICNGVADNLLTNL--AAAR 111
QY 128 HIPIMIAAMHETMYRHPVIVRENIERLKKLVGEFIPRI-----EGRKAVASIDEIVY 181
DB 112 KCPLAIPANVMVWMLNPANQRIALQVSDGITVVMPLGEGQACRENGMRPEPAELLD 171
QY 182 RVIKKLHKTLGKRVLTAGATREYIDPIRFITNASSGKMGVALAEADFRGA-VTLIR 240
DB 172 LLPDLWTTPKILDKKILITAGATFEAIDPVRGITNTSSGKMGVALACRACAGAEISLTH 231
QY 241 TKGSKVAFRIKIKLVETV--EEMLSAIEENELRSKKYDVVIMAAVSDFRPKIAEGKI 298
DB 232 --GQLOTTLPFGISITVQVNSAEDMRAVHRLI--EKQDAFISVAASVDYKVKNRSTQKF 287
QY 299 ---KSGRSITTELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEEGKQIERAKADL 355
DB 288 KKDKNAPLSIEL-DENPDTLASIASLPNPPFCIGFAET--ENVMAVAREKRIKKKIPV 344
QY 356 VVGNTLE-AFGSEENOVVLIGRDFTKELPKMKKRELAERIWEIEKXLS 403
DB 345 IVANDVSIAMGKTTNQIIVIDDADLSFPETSKDEAAMRIVERLAVLYN 393
```



```
XX SQ Sequence 401 AA;
Query Match 22.2%; Score 444; DB 6; Length 401;
Best Local Similarity 32.7%; Pred. No. 4.9e-34;
Matches 137; Conservative 81; Mismatches 145; Indels 56; Gaps 18;
QY 16 LVGKIVXXPGSIAAL-DVKACEGLIRHGAHVHVMSEAATKIHPVAMNLTGNPVIT 74
DB 2 LKNKKIALYVTGGIAVYKSLYLLREIIKQGGVVRVMTQAACQFVNPISFQVLSQKV-- 59
QY 75 EITGF-----IEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTAPP 127
DB 60 QIDTFEEGQPEVSVIDLT---DWADYSIVAPATANIIGKLANGIGDDFVSTAL-LATD 114
QY 128 HIPMIAPAMHETMYRHPVIVRNIERLKKLGVEFGPRI-----EEGRKVASIDEIV- 180
DB 115 H-PIPLVPAMNTKMYENPALKKNKAFLIEQGHYWMPEDIGFLAEGYEGGLRFPDLDRIMA 173
QY 181 ---YRVIKLHKKTLEGGKRLVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-V 236
DB 174 EFNHFIAR-NPGILSGKKVLVTAGTVERIDPVRIYISNDSSGKMGHQLAQAAYEAGQV 232
QY 237 TLIR-----TKGSVKAFRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPK 291
DB 233 SLVTASDLPTSPFIDRFQVESTLDLYQTVSDLYD-----HHDILMMAAAVSDYRPV 283
QY 292 IKAEGKIKSGRSITIELVPXNPKII---DRIKEIQPNVFLVGFKAETSKEKLIIEGKQOI 348
DB 284 NRSDDKMKKQDNLTIEL-EKNPDILAEMGRRKDQGIN---VGFAAET--HNLEEYAOQKL 337
QY 349 ERAKADLVVGNLT---EAFGSEENQVVLIGRDTFK-ELPKMKKRELAERIWDIEKXL 402
DB 338 ASKQADLIVANEVGRDGRGNADENAAVLFSSDQDPLELPLOSCKDMAKKIIIEVVASKL 396

RESULT 15
ADJ27133
ID ADJ27133 standard; protein; 401 AA.
XX AC ADJ27133;
XX DT 20-MAY-2004 (first entry)
XX DE Alloicoccus otitidis protein, a novel antibacterial target SeqID 78.
XX KW growth and survival; otitis media with effusion; OME;
XX KW bacterial infection; antibacterial; antiinfective.
XX OS Alloicoccus otitis.
XX PN WO2003104391-A2.
XX PD 18-DEC-2003.
XX PF 25-NOV-2002; 2002WO-US036122.
XX PR 29-NOV-2001; 2001US-0333777P.
XX PR 18-NOV-2002; 2002US-0426742P.
XX PR 18-NOV-2002; 2002US-0427367P.
XX XX (AMHP ) WYETH HOLDINGS CORP.
XX PI Murphy E, Projan SJ;
XX DR WPI; 2004-053616/05.
XX DR N-PSDB; ADJ27132.
XX PT Novel isolated Alloicoccus otitidis polypeptides useful for identifying
XX PT compounds that inhibit the activity or expression of the polypeptide and
XX PT thus are useful for treating infections caused by Alloicoccus otitidis.
XX PS Claim 7; SEQ ID NO 78; 433pp; English.
```

```
XX CC This invention relates to novel isolated Alloicoccus otitidis (A.
CC otitidis) nucleic acid molecules and encoded proteins thereof.
CC Specifically, it refers to proteins that are essential for the growth and
CC survival of the gram-positive bacterium A. otitidis, and hence provide
CC novel antibacterial targets. The present invention describes
CC pharmaceutical compositions and antisense compounds that are useful for
CC inhibiting activity or expression of these proteins. Furthermore, it
CC provides diagnostics and therapeutics that can be used to ameliorate
CC diseases that are associated with A. otitidis, such as otitis media with
CC effusion (OME) and various bacterial infections. Accordingly, these
CC compositions exhibit both antibacterial and antiinfective activities.
CC This polypeptide is an A. otitidis protein sequence of the invention.
XX SQ Sequence 401 AA;
Query Match 22.2%; Score 444; DB 8; Length 401;
Best Local Similarity 32.7%; Pred. No. 4.9e-34;
Matches 137; Conservative 81; Mismatches 145; Indels 56; Gaps 18;
QY 16 LVGKIVXXPGSIAAL-DVKACEGLIRHGAHVHVMSEAATKIHPVAMNLTGNPVIT 74
DB 2 LKNKKIALYVTGGIAVYKSLYLLREIIKQGGVVRVMTQAACQFVNPISFQVLSQKV-- 59
QY 75 EITGF-----IEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTAPP 127
DB 60 QIDTFEEGQPEVSVIDLT---DWADYSIVAPATANIIGKLANGIGDDFVSTAL-LATD 114
QY 128 HIPMIAPAMHETMYRHPVIVRNIERLKKLGVEFGPRI-----EEGRKVASIDEIV- 180
DB 115 H-PIPLVPAMNTKMYENPALKKNKAFLIEQGHYWMPEDIGFLAEGYEGGLRFPDLDRIMA 173
QY 181 ---YRVIKLHKKTLEGGKRLVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-V 236
DB 174 EFNHFIAR-NPGILSGKKVLVTAGTVERIDPVRIYISNDSSGKMGHQLAQAAYEAGQV 232
QY 237 TLIR-----TKGSVKAFRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPK 291
DB 233 SLVTASDLPTSPFIDRFQVESTLDLYQTVSDLYD-----HHDILMMAAAVSDYRPV 283
QY 292 IKAEGKIKSGRSITIELVPXNPKII---DRIKEIQPNVFLVGFKAETSKEKLIIEGKQOI 348
DB 284 NRSDDKMKKQDNLTIEL-EKNPDILAEMGRRKDQGIN---VGFAAET--HNLEEYAOQKL 337
QY 349 ERAKADLVVGNLT---EAFGSEENQVVLIGRDTFK-ELPKMKKRELAERIWDIEKXL 402
DB 338 ASKQADLIVANEVGRDGRGNADENAAVLFSSDQDPLELPLOSCKDMAKKIIIEVVASKL 396
```

Search completed: July 2, 2005, 12:22:41
Job time : 89.0698 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 2, 2005, 08:53:17 ; Search time 2821 Seconds
(without alignments)
2104.937 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKELIIPFSEE.....PYRGNYQGSTRLAFSKRKL 156

Scoring table: BLOSUM62
-Q= gapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cn2 1/USPTO spool/US08957709/runat 01072005 154632 26016/app query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US08957709 @CGN 1 1 4352 @runat 01072005 154632 26016 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: gb_est1: *
2: gb_est2: *
3: gb_hc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gsa1: *
9: gb_gsa2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	23.8	600	BZ893227	HL12_0182
2	162.5	20.2	792	CL680050	PR10127C
3	150	18.6	1089	BZ551941	pac61-60
4	125	15.5	465	CD084073	MA3-9999SU
5	124	15.4	586	CB336913	Tc024E08F
6	122	15.1	491	A1395397	MA003489
7	122	15.1	507	CD145288	MGI-0104U
8	122	15.1	530	CB335318	Tc005E10R
9	119	14.8	800	CK315881	SB02028A1

10	119	14.8	826	7	CK311070	CK311070 SB02007B2
11	118.5	14.7	519	1	AA968575	oa75508.8
12	118.5	14.7	839	8	BZ549391	pac61-60
13	117	14.5	663	1	AJ451538	AJ451538
14	116.5	14.5	481	7	CO297884	EK170827
15	116	14.4	444	7	CV037326	4140391.B
16	116	14.4	474	6	CB018433	pgn1c.pk0
17	116	14.4	519	6	CB110414	K-EST0151
18	116	14.4	587	5	BX336627	603497875
19	116	14.4	597	5	BX262764	BX262764
20	116	14.4	597	5	BX262765	BX262765
21	116	14.4	606	5	BX304266	603608777
22	116	14.4	608	1	AJ445450	AJ445450
23	116	14.4	632	6	CD216556	pgm2n.pk0
24	116	14.4	636	5	BUI04949	603005879
25	116	14.4	642	4	BM486691	pgm2n.pk0
26	116	14.4	648	5	BX395616	603805513
27	116	14.4	653	5	BX390007	603512155
28	116	14.4	664	6	CD218033	pgn1n.pk0
29	116	14.4	676	5	BUI12250	603157230
30	116	14.4	686	5	BUI296179	603609974
31	116	14.4	701	7	CO422270	GGEZHT101
32	116	14.4	714	6	CD762467	GGEZSM102
33	116	14.4	801	5	BUI421023	603234106
34	116	14.4	829	5	BUI13346	603155084
35	116	14.4	830	5	BUI27586	603112856
36	116	14.4	842	5	BUI317711	603487334
37	116	14.4	925	5	BUI08606	603110802
38	116	14.4	937	7	CO648549	ILLUMIGEN
39	116	14.4	1066	5	BUI473281	603364912
40	115.5	14.3	521	4	B1236417	RE32469.5
41	115.5	14.3	522	3	AY231831	AY231831 Drosophila
42	115.5	14.3	523	4	B1233284	B1233284 RE29587.5
43	115.5	14.3	534	4	B1236440	B1236440 RE32493.5
44	115.5	14.3	545	1	A1512510	A1512510 LD44405.5
45	115.5	14.3	548	4	B1173237	B1173237 RE16016.5

ALIGNMENTS

RESULT 1
BZ893227
LOCUS HL12_0182 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
DEFINITION genomic survey sequence.
ACCESSION BZ893227.1 GI:33343817
VERSION BZ893227.1
KEYWORDS GSS.
SOURCE Halorubrum lacusprofundi
ORGANISM Halorubrum lacusprofundi
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halorubrum.
REFERENCE 1 (bases 1 to 600)
AUTHORS Goo,Y., Roach,J., Glusman,G., Baliga,N.S., Deutsch,K., Pan,M., DasSarma,S., Ng,W.V. and Hood,L.
TITLE Low-pass Sequencing for Microbial Comparative Genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoo@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun.
Location/Qualifiers
1..600
/organism="Halorubrum lacusprofundi"
/mol_type="genomic DNA"
/strain="ATCC 49239"
/db_xref="taxon:2247"
/clone_lib="H1 pUC18 Library"
/note="Vector: pUC18; Site_1: SmaI; A shotgun library was

constructed from Halorubrum lacusprofundi genomic DNA
using pUC18/SmaI/BAP plasmid"

ORIGIN

Alignment Scores:
Pred. No.: 5e-14 Length: 600
Score: 192.00 Matches: 55
Percent Similarity: 50.57% Conservative: 34
Best Local Similarity: 31.25% Mismatches: 53
Query Match: 23.82% Indels: 34
DB: 8 Gaps: 8

US-08-957-709A-71 (1-156) x BZ893227 (1-600)

```
Qy 1 MetLeuLeuProAspTyrLysIle-----ArgLysGluLeuLeuGlu 15
|||||
Db 31 ATGATCTCTCGGACCGGACATCTCGCGCTCGCGAGGGGACCTCGTCGTGAA 90
|||
Qy 16 ProPheSer-----GluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg 33
|||||
Db 91 CCGCTCGACGACGTCGACACGAGGTCAGCGCGGCGGCGTGGAGCTGGCGTGGC 147
|||
Qy 34 GluAlaPheValLys-----Gly 39
|||||
Db 148 GAGCGCTCTCGGAGTTCAGCGCTTTATCTCTGTATCCACCCAAACGAGCGCGAC 207
|||
Qy 40 LysLeuIleAspValGluLysGluLysVal-----ValIlePro 53
|||||
Db 208 GAGTCTCGCGACTACGTCCAGGACGCGCTCCCGAGGCGGAGGATTCATCTCCAC 267
|||
Qy 54 ProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspValMet 73
|||||
Db 268 CCGGCGGACTCGCTCGGACACACGAGCGGCTCGAGATCCCGCGGACTGCTC 327
|||
Qy 74 GlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIle-----GlySerPhe 91
|||||
Db 328 GCGACCGTGAAGCGCGCTCTCGTGGGCGGCTCGCGATCGTCATCCAGCGCCACGCG 387
|||
Qy 92 AlaTrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsn 111
|||||
Db 388 GGGATCGTCGATCCGGGGTACAAGGGGCGAGATCGCTCGAACTTCGAACTCGGAC 447
|||
Qy 112 GluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGlu 131
|||||
Db 448 GNGCCGCTCGGCTCACCGCGGATGCGCTCTCACAGCTCATCTTCACGAGCTCAA 507
|||
Qy 132 GlyProAlaArgAsnProTyr-----ArgGlyAsn-----TyrGln 143
|||||
Db 508 TCGCCCGCAAAACGCGCTATGGCGTCGACGCGGCTCGAAGTATCAG 555
|||
```

RESULT 2

CL680050/c 792 bp DNA linear GSS 09-JUL-2004
LOCUS PRI0127c.H08.2 - PRI0127c.BR (792) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

ACCESSION CL680050
VERSION CL680050.1 GI:50186880
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus

REFERENCE 1 (bases 1 to 792)
AUTHORS Srinivasan, J., Otto, G.W., Kahlow, U., Geisler, R. and Sommer, R.J.
TITLE AppADB: an AcedB database for the nematode satellite organism
JOURNAL Pristionchus pacificus
COMMENT Nucleic Acids Res. 32 (1), D421-D422 (2004)

CONTACT: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel.: 00497071601371

Fax: 00497071601498

Email: ralf.sommer@uebingen.mpg.de

This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.

Seq primer: T7

Class: fosmid ends.

FEATURES

source

Location/Qualifiers
1..792
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN

Alignment Scores:
Pred. No.: 4.69e-10 Length: 792
Score: 162.50 Matches: 41
Percent Similarity: 50.93% Conservative: 14
Best Local Similarity: 37.96% Mismatches: 44
Query Match: 20.16% Indels: 9
DB: 9 Gaps: 2

US-08-957-709A-71 (1-156) x CL680050 (1-792)

```
Qy 54 ProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspValMet 73
|||||
Db 672 CCAGGAGAGTGGCGGTGCGGTGACGTGAGCTGGCGTGGCGTGGCGTGGCGTGG 613
|||
Qy 74 GlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
|||||
Db 612 GGCTGGCTGACGCGGCTTCTCTTTGGCGGTCTGGGCTGATGGTGCACGTCACCG 553
|||
Qy 93 ---TrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsn 111
|||||
Db 552 CACCGTATCGATCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 493
|||
Qy 112 GluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGlu 131
|||||
Db 492 CTGCGCTGGCGTGGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 433
|||
Qy 132 GlyProAlaArgAsnProTyrArg-----GlyAsnTyrGlnGly 144
|||||
Db 432 GGCGCGGCGCGCGACCTTACACCGCGGTGAGATGCGAATATCGCAACACGAGG 373
|||
Qy 145 SerThrArgLeuAlaPheSerLys 152
|||
Db 372 GCGGTAGCCAGCGCAATCGATAAA 349
|||
```

RESULT 3

BZ551941

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ551941 1089 bp DNA linear GSS 17-DEC-2002
pacsl-60.3403.x1 pacsl-60 Pseudomonas aeruginosa genomic clone
pacsl-60.3403, genomic survey sequence.

BZ551941

BZ551941.1 GI:27155771

GSS.

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 1089)

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES

source
Location/Qualifiers
1..1089
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacel-60 3403"
/clone_lib="pacel-60"
/note="clinical isolate 1-60 Whole genomic shotgun library."

ORIGIN

Alignment Scores:
Pred. No.: 2,99e-08 Length: 1089
Score: 150.00 Matches: 37
Percent Similarity: 47.86% Conservative: 30
Best Local Similarity: 26.43% Mismatches: 55
Query Match: 18.61% Indels: 18
DB: 8 Gaps: 3

US-08-957-709A-71 (1-156) x BZ551941 (1-1089)

Qy 30 ArgValGlyArgGluAlaPheValLysGlyLys----- 40
Db 25 CGAATTGGTACCGGCCCCCTCGAGTCGACGATCGATAAGCTTGATATCGAATT 84
Qy 41 -----LeuLeuValGluLysGluGlyValValLeuProProArg 55
Db 85 CCTGCAGCCCAAGCTTCGTGCATCAACACGCGAC---GTCTGCATCATCCCGCGAAC 141
Qy 56 GluTyrAlaLeuLeuLeuLeuGluArgGlyLeuLysLeuProAspValMetGlyAsp 75
Db 142 TCCTTCGCCCTCGCGGACCGCTCGAGTACTTCGCGATCCCGCGAGCTCTGACCATC 201
Qy 76 MetLysIleArgSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAsp 95
Db 202 TGCTGGGCAAGCAGCACCTACGCGGTTGCGGCATCATCTCACTCACTCCCGCTGGAG 261
Qy 96 ProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGlu 115
Db 262 CGGAGTGGAGGCGCATGTGACCTCGAGTTCTCCATACACCAACCTCGCGCGAG 321
Qy 116 LeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArg 135
Db 322 ATCTAGCCCAATGAAGCGTGGCGCAGATGCTTCTCGCAATCCGACGAGCGCTCGGAA 381
Qy 136 AsnProTyrArg-----GlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 382 GTGTCTATAGGACCGTGGCGGCAATACGAGGCCGCGCGGTGACCTGCGCAAA 441

RESULT 4

CD084073
LOCUS MA3-9999U-V270-H05-U.B MA3-0001 Schistosoma mansoni cDNA clone
DEFINITION MA3-9999U-V270-H05.B, mRNA sequence.
ACCESSION CD084073
VERSION CD084073.1 GI:34635056
KEYWORDS EST.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidae; Schistosomatidae; Schistosomatidae; Schistosoma.
1 (bases 1 to 465)
Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,

Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Serubal, J.C., Leite, L.C. and Dias-Neto, E.
Transcriptome analysis of the acoelomate human parasite Schistosoma
mansoni
Nat. Genet. 35 (2), 148-157 (2003)
22879926
12973350
Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjovski@iq.usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: MA3-9999U-V270 row: 5 column: H.
Location/Qualifiers
1..465
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MA3-9999U-V270-H05.B"
/sex="mixed pool"
/dev_stage="adult"
/lab_host="Mus musculus"
/clone_lib="MA3-0001"

FEATURES

source

ORIGIN

Alignment Scores:
Pred. No.: 1.53e-05 Length: 465
Score: 125.00 Matches: 41
Percent Similarity: 43.70% Conservative: 18
Best Local Similarity: 30.37% Mismatches: 52
Query Match: 15.51% Indels: 24
DB: 6 Gaps: 3
US-08-957-709A-71 (1-156) x CD084073 (1-465)

Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
Db 54 TCTCTCTTGGCGCTGGTTACGATCTTACGCTGCTCATGAAGCT----- 98
Qy 41 LeuIleAspValGluLysGluLysValValIleProArgGluTyrAlaLeuIle 60
Db 99 -----ACCATACCTCTCGGGGTCGTAATTAAIT 128
Qy 61 LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleArgSer 80
Db 129 AAACTGACATACAGATTGAATTGGCGGAGGCTGTTATGTCGCTCCGCCAAGC 188
Qy 81 SerLeuAla---ArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAsp 99
Db 189 GGACTAGCTTAAACACAGGGGATTGTTGCTGCTGTTGTTATTCACCGACTATAGG 248
Qy 100 GlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGly 119
Db 249 GCGAATGTTGGTGTAGTACTTTTCAACTTTGGAGAACAGAAATTTGAAATCAGAAAGGGT 308
Qy 120 GluArgPheValGlnIleAlaPheIleArgLeu-----Glu 131
Db 309 GACCTGTAGCACTGATTGTTGAACTATTTCTGCTCCTGAATGATGTAATGTGA 368
Qy 132 GlyProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThr 146
Db 369 TCGTTGGCGGACGAGCGATCGTGGTTCAATGTTATGCTCAACT 413
RESULT 5
CB336913/c

LOCUS CB336913 586 bp mRNA linear EST 01-JAN-2004
DEFINITION Tc024E08F Tribolium castaneum embryonic cDNA library Tribolium
castaneum cDNA clone Tc024E08 3', mRNA sequence.

ACCESSION CB336913
VERSION CB336913.1 GI:40544638

KEYWORDS EST.
SOURCE Tribolium castaneum (red flour beetle)

ORGANISM Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
Tenebrionidae; Tribolium.

REFERENCE 1 (bases 1 to 586)
AUTHORS Savard,J. and Tautz,D.

TITLE A Tribolium castaneum EST project
JOURNAL Unpublished (2003)

COMMENT Contact: Savard, J.

Abteilung fur Evolutionsgenetik, AG Tautz

Institut fur Genetik, Universitat zu Koln

Weyertal 121, 50931 Koln, Germany

Tel: 49 221 470 6911

Fax: 49 221 470 5975

Email: savard@uni-koeln.de

Seq primer: M13F -20.

Location/Qualifiers

1..586

/organism="Tribolium castaneum"

/mol_type="mRNA"

/strain="Wild type"

/db_xref="taxon:7070"

/clone="Tc024E08"

/dev_stages="Mixed embryonic stages"

/clone_libs="Tribolium castaneum embryonic cDNA library"

/note="Vector: pBluescript SK; Site 1: EcoRI; site 2:

XhoI; Uni-ZAP XR cDNA library (Stratagene) constructed by

Reinhard Schroder (1995)"

ORIGIN

Alignment Scores:

Pred. No.: 2,85e-05 Length: 586

Score: 124.00 Matches: 43

Percent Similarity: 50.38% Conservative: 24

Best Local Similarity: 32.33% Mismatches: 43

Query Match: 15.38% Indels: 24

DB: 6 Gaps: 4

US-08-957-709A-71 (1-156) x CB336913 (1-586)

Qy 16 PropheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAla 35

Db 496 CCTCCGACCAAGGTTTCAGTTAAAGCCGCGGTATGATCTG-----AAGAGCGCT 446

Qy 36 PheValLysGlyLysLeuLeuAspValGluLysGluGlyLysValValleProProArg 55

Db 445 TTT-----GACTGCGTGGTCCCGCCGA 422

Qy 56 GluTyrAlaLeuLeuThreuGluArgileLysLeuProAspValMetGlyAsp 75

Db 421 GGTAAAGGCCCTCGTGGACCGCCGATCAAAATCCAGCTCCCGAAGGGTGCTACGACGG 362

Qy 76 MetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrp---Val 94

Db 361 ATCGCCCTAGTCTGGGCTAGCAGTCAAGATTTTCATCGATGTGGCGCGGGGTCGTG 302

Qy 95 AppProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProVal 114

Db 301 GATGAGGACTACAGGGGGTCTCTCAAGGTGGTCTGTTCACCAATTCAGATACCTCGTTT 242

Qy 115 GluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAla 134

Db 241 GAAGTCACAGAGTGGCGCGCATTCACAGCTAAATTCGACGCGGATT----- 194

Qy 135 ArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArg 147

Db 193 -----TATTAT-CCCCGATATCGAGGAAGTTCAAGA 165

RESULT 6

LOCUS AI395397/c

DEFINITION MA003489.C8F Soares normalized S8W Schistosoma mansoni cDNA 3',

ACCESSION AI395397

VERSION AI395397.1 GI:4224944

KEYWORDS EST.

SOURCE Schistosoma mansoni

ORGANISM Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

REFERENCE 1 (bases 1 to 491)

AUTHORS Bailey,J.A., Bentley,K., Blanton,R.E., Soares,M.B. and

Chakravarti,A.

TITLE Expressed sequence tags from a normalized adult Schistosoma mansoni

library

JOURNAL Unpublished (1999)

COMMENT Contact: Chakravarti A

Department of Genetics

Case Western Reserve University

2109 Adelbert Rd, Cleveland, OH 44106, USA

Tel: 216-368-5847

Fax: 216-368-5857

Email: axc39@po.cwru.edu

Additional data regarding this EST may be found at

http://genome.cwru.edu/schistosoma/est/S8west.html

Seq primer: M13 Forward.

Location/Qualifiers

1..491

/organism="Schistosoma mansoni"

/mol_type="mRNA"

/strain="Puerto Rican"

/db_xref="taxon:6183"

/sex="mixed"

/dev_stage="8 week old adult worms"

/lab_host="Mus musculus"

/clone_lib="Soares normalized S8W"

/note="Vector: p77T3-pac; Site_1: NotI; Site_2: EcoRI;

Note: normalization has resulted in an increased frequency

of the rare contaminating mouse (host) sequences (~10%).

For details of library construction see: Bonaldo MF,

Lennon G, Soares MB. Normalization and Subtraction: Two

Approaches to Facilitate Gene Discovery. 1996. Genome

Research 6:791-806"

ORIGIN

Alignment Scores:

Pred. No.: 4.03e-05 Length: 491

Score: 122.00 Matches: 41

Percent Similarity: 42.96% Conservative: 17

Best Local Similarity: 30.37% Mismatches: 53

Query Match: 15.14% Indels: 24

DB: 1 Gaps: 3

US-08-957-709A-71 (1-156) x AI395397 (1-491)

Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40

Db 449 TCTCTCTTCGCGCTGGTTACGATCTTTACGCTGCTCATGAAGCT----- 405

Qy 41 LeuIleAspValGluLysGluGlyLysValValleProProArgGluTyrAlaLeuIle 60

Db 404 -----ACCATCTCTCTGGGGGTCGTGAATTAATT 375

Qy 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80

Db 374 AAAACTGACATACAGATTGAATTCGCGAAGGCTGTTATGTCGCGTCGCTCCACGAGC 315

Qy 81 SerLeuAla---ArgGluGlyValIleGlySerPheAlaTrpPheAlaTrpProGlyTyrAsp 99

```

Db      314  GGACTAGCCTTAAACAGGGGATTGATGTGGTGGTGGTGTATTATACCGCGACTATAGG 255
Qy      100  GlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGly 119
Db      254  GGCAATGTTGGTGTAGTACTTTTCAACTTTGGAGAACCCAGAAATTTGAAATCAGAAAGCGT 195
Qy      120  GluArgPheValGlnIleAlaPheIleArgLeu-----Glu 131
Db      194  GACCGGTGACCAACTGATTGTGACGATATTTTCTGCTGAACGTGATGGAATGTGAA 135
Qy      132  GlyProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThr 146
Db      134  TCGTTCGGGAGAGCGATCGTGTTCAAATGGTTATGGCTCAACT 90

RESULT 7
CD145288/c
LOCUS
DEFINITION
MGI-0104U-A343-G06-U.G MGI-0104 Schistosoma mansoni cDNA clone
CD145288
CD145288.1 GI:34683105
EST.
Schistosoma mansoni
Schistosoma mansoni
Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeida; Schistosomatidae; Schistosomatidae; Schistosoma.
1 (bases 1 to 507)
Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni
JOURNAL
MEDLINE
PUBMED
12973350
COMMENT
Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjo@iq.usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: MGI-0104U-A343 row: 6 column: G.
FEATURES
source
1..507
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MGI-0104U-A343-G06.G"
/sex="mixed pool"
/dev_stage="germball"
/lab_host="Biomphalaria glabrata"
/clone_lib="MGI-0104"
/note="Vector: pGEM T-easy"
ORIGIN
Alignment Scores:
Pred. No.: 4,21e-05 Length: 507
Score: 122.00 Matches: 36
Percent Similarity: 47.37% Conservative: 18
Best Local Similarity: 31.56% Mismatches: 44
Query Match: 15.14% Indels: 16

```

```

DB: 6 Gaps: 2
US-08-957-709a-71 (1-156) x CD145288 (1-507)
Qy      21  SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
Db      305  TCTCTCTTTGGCGCTGGTTACGATCTTTACGCTGCTCATGAAGCT-----261
Qy      41  LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAlaLeuIle 60
Db      260  -----ACCATACCTCTCTGGGGGTCTGTAATTAAT 231
Qy      61  LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
Db      230  AAAACTGCATACATGAAATTCGCGAAGCTGTTATGGTCGCGCTCCACCAAGC 171
Qy      81  SerLeuAla---ArgGluGlyValIleGlySerPheAlaTyrValAspProGlyTyrAsp 99
Db      170  GGGCTAGCCTTAAACAGGGGATGATGTTGGTGGTGGTATTGACCGCGACTATAGG 111
Qy      100  GlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGly 119
Db      110  GGCAATGTTGGTGTAGTACTTTTCAACTTTGGAGAACCAAGAAATTCGAATCAGAAAGGT 51
Qy      120  GluArgPheValGlnIleAlaPheIleArgLeuGluGlyPro 133
Db      50  GACCGGTGACCAACTGATTTGTGAACGTATTTTCTGTCTCT 9

RESULT 8
CB335318/c
LOCUS
DEFINITION
Tc005E10R Tribolium castaneum embryonic cDNA library Tribolium
castaneum cDNA clone Tc005E10 5', mRNA sequence.
CB335318
CB335318.1 GI:40543043
EST.
Tribolium castaneum (red flour beetle)
Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
Tenebrionidae; Tribolium.
1 (bases 1 to 530)
Savard,J. and Tautz,D.
A Tribolium castaneum EST project
Unpublished (2003)
Contact: Savard, J.
Abteilung für Evolutionsgenetik, AG Tautz
Institut für Genetik, Universität zu Köln
Weyertal 121, 50931 Köln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Seq primer: ML3R.
FEATURES
source
1..530
/organism="Tribolium castaneum"
/mol_type="mRNA"
/strain="Wild type"
/db_xref="taxon:7070"
/clone="Tc005E10"
/dev_stage="Mixed embryonic stages"
/clone_lib="Tribolium castaneum embryonic cDNA library"
/note="Vector: pBluescript SK; Site 1: EcoRI; Site 2:
XhoI; Uni-ZAP XR cDNA library (Stratagene) constructed by
Reinhard Schroder (1995)"
ORIGIN
Alignment Scores:
Pred. No.: 4,48e-05 Length: 530
Score: 122.00 Matches: 43
Percent Similarity: 49.62% Conservative: 23
Best Local Similarity: 32.33% Mismatches: 44
Query Match: 15.14% Indels: 24

```

```
DB:
US-08-957-709a-71 (1-156) x CB335318 (1-530)
Qy 16 PropheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAla 35
Db 503 CCTCCGACCAAGGTTCTGTTAAAGCCGCGGTTATGATCTG-----AAGACGCGCT 453
Qy 36 PheValLysGlyLysLeuLeuAspValGluLysGluGlyLysValValValValValProArg 55
Db 452 TTT-----GACTGGGTGTCCTCCGCGCCGA 429
Qy 56 GluTyrAlaLeuLeuLeuThrLeuGluArgLleLysLeuProAspAspValMetGlyAsp 75
Db 428 GGTAGGCGCTCGTGACACCGCATCAAAATCCAGCTCCCGAAGGCTGCTACGACGG 369
Qy 76 MetLysIleArgSerSerLeuAlaArgGluGlyValLleGlySerPheAlaTrp---Val 94
Db 368 ATCGCCCTAGGTCGGGTCTAGCAGTCAAGAAATTTTCATCGATGTTGGCGCGGGGTCGTG 309
Qy 95 AspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProVal 114
Db 308 GATGAGGACTACAGGGGGTCTCAAGGTGGTCTATTCAACCATTCAGATACCTCGTTT 249
Qy 115 GluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAla 134
Db 248 GNAGTCCAGATGGGCGACCGCATTCACAGCTAATTTGTGAGCGGATC----- 201
Qy 135 ArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArg 147
Db 200 -----TATTA-CCCGATATCGAGGAGTTCAAGA 172

RESULT 9
CK315881
LOCUS
DEFINITION
SB02028A1F02.f1 normalized Keck-Tagu Library SB02 Taeniopygia
guttata cDNA clone SB02028A1F02.f1 5, mRNA sequence.
ACCESSION
CK315881
VERSION
CK315881.1 GI:44825455
KEYWORDS
EST.
SOURCE
Taeniopygia guttata
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Passeriformes; Estrilidae;
Estrildinae; Taeniopygia.
REFERENCE
1 (bases 1 to 800)
Clayton,D.F., Arnold,A.P., Ball,G.F., Brenowitz,E., George,J.M.,
Mello,C.V., Wade,J., Replogle,K., Lewin,H., Band,M., Hernandez,A.
and Liu,L.
The Songbird Neurogenomics Initiative: An Evolving Public Resource
for Study of Genes, Brain, and Behavior
Unpublished (2004)
JOURNAL
COMMENT
Contact: David F. Clayton
University of Illinois
B107 CILSL, 601 S. Goodwin, Urbana, IL 61801, USA
Tel: 217 244 3668
Fax: 217 244 1648
Email: dclayton@uiuc.edu
Base Calling/Quality Scores: PHRED from Washington University
Genome Center.
Vector Trimming: Cross match from Washington University Genome
Center PHRAP suite. Low quality bases (Phred score < 20) were
trimmed from both ends of the sequence by an in-house script.
This sequence is vector free and at least 200 bp in length. Funded
by PHS grant # RO1 NS045264, 'Songbird Neurogenomics Initiative.'
PCR Primers
FORWARD: TAATACGACTCACTAGG(T7)
BACKWARD: ATTAACCTCACTAAG(T3)
Insert Length: 800 Std Error: 0.00
Plate: SB02028A1 row: F column: 02
Seq primer: TAATACGACTCACTAGG (T7)
High quality sequence stop: 800.
Location/Qualifiers
```

```
source
1. .800
/organism="Taeniopygia guttata"
/mol_type="mRNA"
/db_xref="taxon:59729"
/clone="SB02028A1F02.f1"
/tissue_type="brain"
/dev_stage="late embryo, post-hatch days 1, 10, 20, 45,
and adult (pooled)"
/lab_host="DH10B"
/clone_lib="normalized Keck-Tagu Library SB02"
/note="Organ: brain; Vector: pBS II SK(+); Site_1:
EcoRI(5' side of insert); Site_2: NotI (3' side of
insert); The library was constructed and normalized as
described by Bonaldo, M.F., Lennon, G. and Soares, M.B.
(1996). Genome Research 6(9): 791-806. An identifying tag
was added at the 3'during cDNA synthesis:
insertAAAAAAAAAAAAAAAAATGCGA."

Alignment Scores:
Pred. No.: 0.000194 Length: 800
Score: 119.00 Matches: 37
Percent Similarity: 46.15% Conservative: 17
Best Local Similarity: 31.62% Mismatches: 47
Query Match: 14.76% Indels: 16
DB: 7 Gaps: 2

US-08-957-709a-71 (1-156) x CK315881 (1-800)
Qy 18 SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
Db 115 TCCCGGGGCTCCGCGGCTGCTGCGGGCTACGATCTGTACAGTGGCTATGACTGT----- 168
Qy 38 LysGlyLysLeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyr 57
Db 169 -----GTGATACCAACCATGAGGAAG 189
Qy 58 AlaLeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLys 77
Db 190 GCTGTGGTGAACACAGACATTCATAGCGCTTCTCTGGATGCTATGGCCGAGTAGCA 249
Qy 78 IleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpVal---AspPro 96
Db 250 CCACGTTCTGTTAGCTGCAAGACACTTCATAGACGTTGAGCTGGTGTATTGATGAG 309
Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 310 GATTACAGGGGAAATGTTGGTGTGCTGTCAACTTTGGCAAGGAGACATTTGAAGTT 369
Qy 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyPro 133
Db 370 AAAAAAGGAGATAGAAATGCCAGCTCATCTGTGACGCACTTTATTCCT 420

RESULT 10
CK311070
LOCUS
DEFINITION
SB02007B2G06.f1 normalized Keck-Tagu Library SB02 Taeniopygia
guttata cDNA clone SB02007B2G06.f1 5, mRNA sequence.
ACCESSION
CK311070
VERSION
CK311070.1 GI:44820644
KEYWORDS
EST.
SOURCE
Taeniopygia guttata
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Passeriformes; Estrilidae;
Estrildinae; Taeniopygia.
REFERENCE
1 (bases 1 to 826)
Clayton,D.F., Arnold,A.P., Ball,G.F., Brenowitz,E., George,J.M.,
Mello,C.V., Wade,J., Replogle,K., Lewin,H., Band,M., Hernandez,A.
and Liu,L.
The Songbird Neurogenomics Initiative: An Evolving Public Resource
for Study of Genes, Brain, and Behavior
Unpublished (2004)
JOURNAL
```


COMMENT

Contact: David F. Clayton
University of Illinois
Bi07 CUSL, 601 S. Goodwin, Urbana, IL 61801, USA
Tel: 217 244 3668
Fax: 217 244 1648
Email: dclayton@uiuc.edu
Base Calling/Quality Scores: PHRED from Washington University
Genome Center.

Vector Trimming: Cross_match from Washington University Genome
Center PHRAP suite. Low quality bases (Phred score < 20) were
trimmed from both ends of the sequence by an in-house script.
This sequence is vector free and at least 200 bp in length. Funded
by PHS grant # R01 NS045264, 'Songbird Neurogenomics Initiative.'

PCR Primers

FORWARD: TAATACGACTCATATAGGG (T7)

BACKWARD: ATTACCCCTCAATAAG (T3)

Insert Length: 826 Std Error: 0.00

Plate: SB02007B2 row: G column: 06

Seq primer: TAATACGACTCATATAGGG (T7)

High quality sequence stop: 826.

FEATURES

source

Location/Qualifiers

1..826
/organism="Taeniopygia guttata"

/mol_type="mRNA"

/db_xref="taxon:59729"

/clone="SB02007B2G06.F1"

/tissue_type="brain"

/dev_stage="late embryo, post-hatch days 1, 10, 20, 45,

and adult (pooled)"

/lab_host="DH108"

/clone_lib="normalized Keck-Tagu Library SB02"

/note="Organ: brain; Vector: pBS II SK(+); Site: 1;

ECORI(5' side of insert); Site 2: NotI (3' side of

insert); The library was constructed and normalized as

described by Bonaldo, M.F., Lennon, G. and Soares, M.B.

(1996). Genome Research 6(9): 791-806. An identifying tag

was added at the 3' during cDNA synthesis:

InsertAAAAAAAAAAAAAAAAAATGCCGA."

ORIGIN

Alignment Scores:

Pred. No.:	0.000203	Length:	826
Score:	119.00	Matches:	37
Percent Similarity:	46.15%	Conservative:	17
Best Local Similarity:	31.62%	Mismatches:	47
Query Match:	14.76%	Indels:	16
DB:	7	Gaps:	2

US-08-957-709a-71 (1-156) x CK311070 (1-826)

QY	18	SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal	37
Db	172	TCGCGGGCTCCGCCGCTGCTGGGGCTACGATCTGTACAGTGCCTTACTGCTGT	225
QY	38	LysGlyLysLeuLeuAspValGluLysGluGlyLysValLysValLysProArgGluTyr	57
Db	226	-----GTGATACCAACCCATGGAGAG	246
QY	58	AlaLeuLeuLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLys	77
Db	247	GCTGTGGTGAACACAGACATTCATAGTCTTCTTGGATGCTATGCCGAGTAGCA	306
QY	78	IleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaIrpVal	96
Db	307	CCACGTTCTGTTAGCTCCAAAGCATTCATAGATGTTGGAGCTGTTATTGATGAG	366
QY	97	GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu	116
Db	367	GATTACAGGGAATGTTGGTGTGTACTGTTCACTTTGGCAAGGAGACGTTTGAAGTT	426
QY	117	ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyPro	133
Db	427	AAAAAGGAGATAGAATTCGCCGCTCATCTGTGAACGCATTATTATTCCT	477

RESULT 11

AA968575/c

LOCUS

DEFINITION

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

```
Db 317 CCTTCTGGGTGTATGAAGAGTGCTCCACGGTCAGCGTTGGCTGCAAAACACTTTATT 258
Qy 89 GlySerPheAlaTrpVal---AspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyr 107
Db 257 GATGTAGAGCTGGTGTATAGATGAAGATTATAGAGGAATGTGTGTGTACTGTTT 198
Qy 108 AsnAlaSerAsnGluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPhe 127
Db 197 AATTTCGAAGAAGAAAGTTTGAAGTCAAAAAGGTGATCGCAATTGCACAGCTCATTTGC 138
Qy 128 IleArgLeuGluGlyPro 133
Db 137 GAACGGATTTTATCCA 120

RESULT 12
BZ549391/c
LOCUS BZ549391 839 bp DNA linear GSS 17-DEC-2002
DEFINITION pacsl-60_2014.xl pacsl-60 Pseudomonas aeruginosa genomic clone
pacsl-60_2014, genomic survey sequence.
ACCESSION BZ549391
VERSION BZ549391.1 GI:27152972
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 839)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
Location/Qualifiers
1..839
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsl-60_2014"
/clone_lib="pacsl-60"
/note="clinical isolate 1-60 Whole genomic shotgun
library."
ORIGIN
Alignment Scores:
Pred. No.: 0.00024 Length: 839
Score: 118.50 Matches: 31
Percent Similarity: 45.76% Conservative: 23
Best Local Similarity: 26.27% Mismatches: 35
Query Match: 14.70% Indels: 29
DB: 8 Gaps: 3

US-08-957-709A-71 (1-156) x BZ549391 (1-839)
Qy 13 LeuIleGluProPheSerGluGluSerLeuGlnProAla----- 25
Db 394 ATGATCGAGCGGTCGTGCGAGCGCGAGGTGCGCGCGGACGACGACGCGGGTGATTCC 335
Qy 26 -----GlyTyrAspLeuArgValGlyArgGlu----- 34
Db 334 TACGGGGTGTCCAGCTACGGCTACGACGTGCGCTGCGCGCGCGAATTCAGGTGTTACC 275
Qy 35 -----AlaPheValLysGlyLysLeuIleAspVal 44
Db 274 AACATCCATTCCGGCGGTGTCGATCCGAAGAACTTCGACGAGAAAGCTTCGTCGACATC 215
```

```
Qy 45 GluLysGluGlyLysValValIleProProArgGluTyrAlaLeuIleLeuThrLeuGlu 64
Db 214 AACAGCGAC---GTCGTCATCATCCCGCGAAGACTCTTGGCCCTGGCGCGCACCGTCGAG 158
Qy 65 ArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeuAlaArg 84
Db 157 TACTTCGCGATCCCGCGCGAGCTCTGCACCATCTGCTGGCGCAAGAGCACCTACGCGGT 98
Qy 85 GluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGlyAsnLeu 102
Db 97 TCGCGCATCATCTCAACAGTCACCCCGCTGAGCGCGAGTGGGAAGGCCATGTA 44

RESULT 13
AJ451538
LOCUS AJ451538 663 bp mRNA linear EST 22-APR-2002
DEFINITION AJ451538 riken1 Gallus gallus cDNA clone 29b24r1, mRNA sequence.
AJ451538
ACCESSION AJ451538
VERSION AJ451538.1 GI:20261634
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE 1 (bases 1 to 663)
AUTHORS Buerstedde, J.M.
TITLE Gallus gallus bursal lymphocyte EST
JOURNAL Unpublished (2002)
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institut
Martinstr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
Location/Qualifiers
1..663
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="29b24r1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/clone_lib="riken1"
/note="CB inbred strain"
ORIGIN
Alignment Scores:
Pred. No.: 0.00027 Length: 663
Score: 117.00 Matches: 37
Percent Similarity: 47.01% Conservative: 18
Best Local Similarity: 31.62% Mismatches: 46
Query Match: 14.52% Indels: 16
DB: 1 Gaps: 2

US-08-957-709A-71 (1-156) x AJ451538 (1-663)
Qy 18 SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
Db 136 TCCAAGGCTCTGCACGGCGCGGCTACACTTGTACAGTGCCTATGAC----- 186
Qy 38 LysGlyLysLeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyr 57
Db 187 -----TACGTGATCCCAACCATGGAACAG 210
Qy 58 AlaLeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLys 77
Db 211 GCTGTAGTGAACACAGACATTTCAGATTGCATCTCTGCTGGATGCTATGGCGGATGACA 270
Qy 78 IleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpVal---AspPro 96
Db 271 CCGGTTCTGTTAGTCTGCAAGCACCTTCATAGATGTTGGTGTGGTATTTCATGATGAG 330
Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
```

```

Db      331 GATTACAGGGAATGTTGGGTGGTACTTCAACTTGGCAGAGAGACTTTTGAAGTT 390
      ... |||||... : : : : : |||||... |||||... |||||... |||||...
Qy      117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyPro 133
      ... |||||... |||||... |||||... |||||... |||||... |||||...
Db      391 AAGAAAGAGATAGATGATCCCGAGCTCATCTGTGAACGCACTTTCTATCCT 441
      ... |||||... |||||... |||||... |||||... |||||... |||||...

RESULT 14
CO297884
LOCUS   EK170827.5prme Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE  Drosophila melanogaster (fruit fly)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 481)
AUTHORS  Koczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
  Peterson,E., and Swimmer,C.
TITLE    Exelixis FlyTag EST Project CK01 Library
JOURNAL  Unpublished (2004)
COMMENT  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  Plate: EK1708 row: C column: 3
  High quality sequence stop: 475.
  Location/Qualifiers
    1..481
    /organism="Drosophila melanogaster"
    /mol_type="mRNA"
    /db_xref="taxon:7227"
    /clone="EK170827"
    /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
    /note="Organ: mixed stage embryos, imaginal disks, and
    adult heads; Vector: pCDNA-SK+; Site:1: NotI; Site:2:
    XhoI; Random primed, normalized library from mixed stage
    embryos, imaginal disks, and adult heads."

FEATURES
    source
    1..481
    /organism="Gallus gallus"
    /mol_type="mRNA"
    /strain="Leghorn and broiler"
    /db_xref="taxon:9031"
    /clone="3GAL 20E01"
    /lab_host="DH5alpha"
    /clone_lib="BARC 3GAL chicken mixed tissue"
    /note="Vector: pBluescript, SK+, Stratagene; Site:1: NotI;
    Site:2: EcoRI; Normalized library of pooled RNA isolated
    from whole brain, ultimobranchial gland, parathyroid
    gland, cecal tonsil and primordial germ cells Multiple"

ORIGIN
Alignment Scores:
Pred. No.: 0.0002 Length: 481
Score: 116.50 Matches: 35
Percent Similarity: 48.76% Conservative: 24
Best Local Similarity: 28.93% Mismatches: 39
Query Match: 14.45% Indels: 23
DB: 7 Gaps: 3

US-08-957-709A-71 (1-156) x CO297884 (1-481)
Qy      18 SerGluSerLeuGlnPro-----AlaGlyTyrAspLeuArg 30
      ...|||... |||||... |||||... |||||... |||||... |||||...
Db      156 ACCGAGATGCTTTGAGCGGTGAGGGATCGCCAAAGCGCGGAGTTGACCTGCGC 215
      ...|||... |||||... |||||... |||||... |||||... |||||...
Qy      31 ValGlyArgGluAlaPheValLysGlyLysLeuLeuAspValGluGlyLysVal 50
      ...|||... |||||... |||||... |||||... |||||... |||||...
Db      216 AGCGCTACGAC-----GTT 230
      ...|||... |||||... |||||... |||||... |||||... |||||...
Qy      51 ValIleProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAsp 70
      ...|||... |||||... |||||... |||||... |||||... |||||...
Db      231 GTGGTCCGCGCAGCGGAAAGCCATTGTCAAGACCGCATCTGCAAGTGCAGGTTCGGAG 290
      ...|||... |||||... |||||... |||||... |||||... |||||...
Qy      71 AspValMetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySer 90
      ...|||... |||||... |||||... |||||... |||||... |||||...
Db      291 GGCTCTACGACGCGCGCCACCATCGGGCTGGCGATGAAGAAGAACTTCATTGATG 350
      ...|||... |||||... |||||... |||||... |||||... |||||...
Qy      91 PheAlaTrp---ValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAla 109
      ...|||... |||||... |||||... |||||... |||||... |||||...

```

```

Db      351 GCGCGCGGTGTGTGGACGAGATTATCGCGCAATCTCGGCTGCTCTGTTCAATCAC 410
      ...|||... |||||... |||||... |||||... |||||... |||||...
Qy      110 SerAsnGluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArg 129
      ...|||... |||||... |||||... |||||... |||||... |||||...
Db      411 TCAGATGTTGATTTCAGAGTGAAGCATGCGACCGCATGCTTCATTGTCGAGCGT 470
      ...|||... |||||... |||||... |||||... |||||... |||||...
Qy      130 Leu 130
      ...|||... |||||... |||||... |||||... |||||... |||||...
Db      471 ATC 473
      ...|||... |||||... |||||... |||||... |||||... |||||...

RESULT 15
COV037326
LOCUS   4140391 BARC 3GAL chicken mixed tissue Gallus gallus cDNA clone
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE  Gallus gallus (chicken)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
  1 (bases 1 to 444)
AUTHORS  Evock-Clover,C.M., Ashwell,C.M., McMurtry,J.P., Lillehoj,H.S.,
  Matukumalli,L.K. and Van Tassel,C.P.
TITLE    Characterization of expressed sequence tags generated from multiple
  chicken tissues
JOURNAL  Unpublished (2004)
COMMENT  Contact: Christina M. Clover
  Growth Biology Laboratory
  Animal and Natural Resources Institute
  Bldg. 200 Rm215 BARC-East, Beltsville, MD 20705, USA
  Tel: 3015048224
  Fax: 3015048623
  Email: chris@anri.barc.usda.gov
  Single pass sequencing. Bases called and trimmed with phred
  0.000925 using options -trim_alt -, -trim_fasta. Vector identified
  by cross match using options -minmatch 12 -minscore 18
  Plate: 20 row: E column: 01
  Seq primer: CCAGTCACGACGTGTGTAACG
  High quality sequence stop: 444.
  Location/Qualifiers
    1..444
    /organism="Gallus gallus"
    /mol_type="mRNA"
    /strain="Leghorn and broiler"
    /db_xref="taxon:9031"
    /clone="3GAL 20E01"
    /lab_host="DH5alpha"
    /clone_lib="BARC 3GAL chicken mixed tissue"
    /note="Vector: pBluescript, SK+, Stratagene; Site:1: NotI;
    Site:2: EcoRI; Normalized library of pooled RNA isolated
    from whole brain, ultimobranchial gland, parathyroid
    gland, cecal tonsil and primordial germ cells Multiple"

FEATURES
    source
    1..444
    /organism="Gallus gallus"
    /mol_type="mRNA"
    /strain="Leghorn and broiler"
    /db_xref="taxon:9031"
    /clone="3GAL 20E01"
    /lab_host="DH5alpha"
    /clone_lib="BARC 3GAL chicken mixed tissue"
    /note="Vector: pBluescript, SK+, Stratagene; Site:1: NotI;
    Site:2: EcoRI; Normalized library of pooled RNA isolated
    from whole brain, ultimobranchial gland, parathyroid
    gland, cecal tonsil and primordial germ cells Multiple"

ORIGIN
Alignment Scores:
Pred. No.: 0.000208 Length: 444
Score: 116.00 Matches: 37
Percent Similarity: 47.01% Conservative: 18
Best Local Similarity: 31.62% Mismatches: 46
Query Match: 14.39% Indels: 16
DB: 7 Gaps: 2

US-08-957-709A-71 (1-156) x COV037326 (1-444)
Qy      18 SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
      ...|||... |||||... |||||... |||||... |||||... |||||...
Db      136 TCACAGGGCTCTGCGACGGCGCGGTACGACTTGTTCAGTGCCTATGAC----- 186
      ...|||... |||||... |||||... |||||... |||||... |||||...
Qy      38 LysGlyLysLeuIleAspValGluLysGluGlyLysValIleProProArgGluTyr 57
      ...|||... |||||... |||||... |||||... |||||... |||||...

```

Tue Jul 5 14:06:23 2005

```
Db 187 -----TACGTGATCCCAACCATGAAAAAG 210
Qy 58 AlaLeuIleLeuThrLeuGluArgIleIysLeuProAspAspValMetGlyAspMetLys 77
Db 211 GCTGTAGTGAACACAGACATTCAGATTCACCTTCCTGCTGGATGCTATGGCCGAGTAGCA 270
Qy 78 IleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpVal---AspPro 96
Db 271 CCGCGTCTCTGTTAGCTGCAAGCAGCACTTCATAGATGTTGGTGGTGTATTATTGATGAG 330
Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 331 GATTACAGGGGAAATGTTGGCGTGGTACTCTTCAACTTTGGCAAGGAGACCTTTGAAAGTT 390
Qy 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyPro 133
Db 391 AAGAAAGGAGATAGAAATTGGCCAGCTCATCTGTGAACGCATTTTCTATCCT 441
```

Search completed: July 2, 2005, 11:06:29
Job time : 2827 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 2, 2005, 10:23:22 ; Search time 524 Seconds
(without alignments)
1867.291 Million cell updates/sec

Title: US-08-957-709a-71
Perfect score: 806
Sequence: 1 MLLPWKIRKEILIEPFSEE.....PYRGNVQGSRLAFSRKKL 156

Scoring table:

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US08957709/runat_01072005_154633_26120/app_query.fasta_1.327
-DB=Published Applications NA -QWTF=fastap -SUPP=us-rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USRR=US08957709 @CGN 1.1 740 @runat_01072005_154633_26120
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	214	26.6	567	9	US-09-738-626-3140
2	214	26.6	3309400	9	US-09-738-626-1
3	210	26.1	573	15	US-10-156-761-4448
4	210	26.1	9025608	15	US-10-156-761-1
5	194	24.1	423	9	US-09-974-300-5011
6	191.5	23.8	2256646	19	US-10-470-565-1
7	174	21.6	640681	9	US-09-790-988-1
8	168.5	20.9	89047	18	US-10-672-787-34
9	164.5	20.4	2731748	19	US-10-297-465A-1
10	160.5	19.9	2964	20	US-10-784-986-3
11	154	19.1	1230025	17	US-10-289-762-1
12	149.5	18.5	586	19	US-10-437-963-14349
13	149.5	18.5	1830121	17	US-10-329-670-1
14	149.5	18.5	1830121	20	US-10-158-865-1
15	149.5	18.5	1830121	22	US-10-981-687-1
16	145.5	18.1	3129	17	US-10-282-132A-13918
17	134.5	16.7	534	10	US-09-975-719-246
18	134.5	16.7	42235	10	US-09-975-719-1
19	132.5	16.4	501	10	US-09-975-719-244
20	125.5	15.6	4658	21	US-10-258-089-16
21	125.5	15.6	4658	21	US-10-258-089-27
22	125.5	15.6	10112	14	US-10-239-804-12
23	125.5	15.6	10112	14	US-10-233-804-70
24	125.5	15.6	10112	15	US-10-134-643-6
25	125.5	15.6	10112	20	US-10-838-906-24
26	125.5	15.6	10112	21	US-10-258-089-5
27	125.5	15.6	10114	14	US-10-239-804-13
28	125.5	15.6	10114	14	US-10-239-804-71
29	125.5	15.6	10114	20	US-10-838-906-25
30	125.5	15.6	10127	21	US-10-258-089-18
31	125.5	15.6	10221	21	US-10-258-089-6
32	125.5	15.6	10392	21	US-10-258-089-17
33	125.5	15.6	10815	15	US-10-134-643-10
34	125.5	15.6	10815	21	US-10-258-089-7
35	125.5	15.6	10930	21	US-10-258-089-8
36	125.5	15.6	11874	15	US-10-134-643-1
37	125.5	15.6	12481	14	US-10-239-804-59
38	125.5	15.6	12481	14	US-10-239-804-69
39	125.5	15.6	12481	15	US-10-134-643-9
40	125.5	15.6	12481	20	US-10-838-906-22
41	125.5	15.6	12481	21	US-10-258-089-12
42	121	15.0	441	21	US-10-472-928-4751
43	121	15.0	477	22	US-10-617-320-924
44	121	15.0	2162598	21	US-10-472-928-4979
45	117	14.5	10223	8	US-08-961-527-73

ALIGNMENTS

RESULT 1

US-09-738-626-3140
; Sequence 3140, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOHI, AKIHIRO
; APPLICANT: IKEDA, NASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```

/ FILE REFERENCE: 249-125
/ CURRENT APPLICATION NUMBER: US/09/738,626
/ CURRENT FILING DATE: 2000-12-18
/ PRIOR APPLICATION NUMBER: JP 99/377484
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: JP 00/159162
/ PRIOR FILING DATE: 2000-04-07
/ PRIOR APPLICATION NUMBER: JP 00/280988
/ PRIOR FILING DATE: 2000-08-03
/ NUMBER OF SEQ ID NOS: 7059
/ SOFTWARE: PatentIn ver. 3.0
/ SEQ ID NO 3140
/ LENGTH: 567
/ TYPE: DNA
/ ORGANISM: Corynebacterium glutamicum
US-09-738-626-3140

```

Alignment Scores:		
Pred. No.:	8.32E-20	567
Score:	214.00	56
Percent Similarity:	50.30%	Conservative: 29
Best Local Similarity:	33.14%	Mismatches: 60
Query Match:	26.55%	Indels: 24
DB:	9	Gaps: 4

US-08-957-709A-71 (1-156) x US-09-738-626-3140 (1-567)

Qy	1	MetLeuLeuProAspTrpIysIleArgLys-----GluIleLeuIleGlu 15
Db	1	GTGCTCTTTTCAGATCGTGACATTCGTAATAACAAATGACGACGGCGACTTTGGGAATTGAA 60
Qy	16	ProPheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
Db	61	CTTTTCGACGCTGAGCTGATTCAGCCGTGCGATGTCGATGTCGCATGGACCGCTACTTTC 120
Qy	34	GluAlaPheValLysGlyLys----- 40
Db	121	CGGGTTTTCATAACTCTAAGTATACACCCACATTTGACCCCTAAGTTGAATCAGGATGAGCTG 180
Qy	41	-----LeuIleAspValGluLysGluGlyLysValLysValLysProArgGluTyrAla 58
Db	181	ACCAGCTTGTTCAGGTTGAGGACCGCGGAGGCTTTGTGCTGCATCCGGGTGAGTTTGTG 240
Qy	59	LeuIleLeuThrLeuGluArgIleLysLysLeuProAspValMetGlyAspMetLysIle 78
Db	241	CTGGCGTCCACGCTGGAAAAGTTCACTTTGGCTGCGCATCTGGCTGGTCTGTTTGGAGGGT 300
Qy	79	ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
Db	301	AGTCTGCTCTTGGTGGCTTGGCTTGTGTGACGCACTCTACTGCTGGTTCATTGATCCT 360
Qy	97	GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db	361	GGTTTTAGTGGTTACATCAGCTTGAGTTGTCCAATGTGGCTAATCTGCGCGATCACGTTG 420
Qy	117	ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
Db	421	TGGCCGGGTATGAAGTGGGGCAGTTGGCTTGTTCAGATGAGTTCCCTTCGGGAGACT 480
Qy	137	ProTyrArgGlyAsnTyrGlnGlySer 145
Db	481	CCGTATGGTTCCGGCAAGCTTCGTTTCG 507

RESULT 2

US-09-738-626-1/C
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SSIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: KEIKO, KEIKO

```

; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Alignment Scores:
Pred. No.: 1,29e-14 Length: 3309400
Score: 214.00 Matches: 56
Percent Similarity: 50.30% Conservative: 29
Best Local Similarity: 33.14% Mismatches: 60
Query Match: 26.55% Indels: 24
DB: Gaps: 4

US-08-957-709A-71 (1-156) x US-09-738-626-1 (1-3309400)
Qy 1 MetLeuProAspTrpLysIleArgLys-----GlutLeuIleGlu 15
Db 3037411 GTGCTTCTTTCAGATCGTCACATTCGTAATCAATTGACCGCAGCGACTGGGAATTGAA 3037352

Qy 16 ProPheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
Db 3037351 CCITTCACGCTGAGCTGATTGACCGCTGAGAGTGTGCGATGTCGCGATGACCGCTACTTTC 3037292

Qy 34 GluAlaPheValLysGlyLys----- 40
Db 3037291 CGGGTTTTCAAATAACTCTAAGTACACCCACATTGACCCCTAAGTTGAATCAGATGAGCTG 3037232

Qy 41 -----LeuIleAspValGluLysGluGlyLysValValIleProArgGluTyrAla 58
Db 3037231 ACCAGCCTTGTCAGGTTGAGGACCGCGAGGGGCTTTGTGTCATCCGGGTGAGTTGTG 3037172

Qy 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIle 78
Db 3037171 CTGCGCTCCACGCTGAAAGATTTCATTTTGCCTCGCGCATCTGGCTGGTCTTTGGAGGGT 3037112

Qy 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
Db 3037111 AAGTCGTCCTTGGTCGCTCTTGGCTTTGTCAGCAGCTACTGCTGCTGTTTCATTGATCCT 3037052

Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 3037051 GGGTTTAGTGGTTACATCAGCTTGGAGATTGTCCAAATGTGGCTAATTCGCCGATCACGCTTG 3036992

Qy 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
Db 3036991 TGCCCGGGATGAAAGTGGGGCAGTTGGCTTTGTTCCAGATGATGTTCCCTCTGCGGAGACT 3036932

Qy 137 ProTyrArgGlyAsnTyrGlnGlySer 145
Db 3036931 CCCTATGTTTCCGGCAAGCTTGGTTGCG 3036905

```

RESIN.T 3

US-10-156-761-4448
; Sequence 4448, Application US/10156761
; Publication No. US20030119018A1

```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4448
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-10-156-761-448

Alignment Scores:
Pred. No.: 3,06e-19 Length: 573
Score: 210.00 Matches: 53
Percent Similarity: 50.30% Conservative: 32
Best Local Similarity: 31.36% Mismatches: 60
Query Match: 26.05% Indels: 24
DB: 15 Gaps: 3

US-08-957-709a-71 (1-156) x US-10-156-761-4448 (1-573)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluLe-----LeuIleGlu 15
Db 1 GTGCTTCTCTCAGACAAGGACATCCGGCCGAGATCGACGCCGAGCGGTTCGATATCGAC 60
Qy 16 ProPheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
Db 61 CCGTACGACGAATCCATCGGTGTCAGCCCTCGAGCATCGATGTCGGCTCGACCGCTACTTC 120
Qy 34 -----GluAlaPheValLys 38
Db 121 CGGGTGTTCGAGAATCACCGCTATCCGATATCGACCCCTTCCTCGAGCAGCGCGATCTG 180
Qy 39 GlyLysLeuIleAspValGluLysGluGlyLysValIleProProArgGluTyrAla 58
Db 181 ACGGCTCTGTCGAGCCCGGAGGCGAGCGCCGCTTCATCTTCATCCGGGGAGTTCGTC 240
Qy 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIle 78
Db 241 CTCGCCAGTACCTACGAGGTCATCTCCCTCCCGAGCAGCTCGCGTCCGGCTGGAGGGG 300
Qy 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
Db 301 AAGAGTTCCTCCGGGGCTCGGGCTGGTGGTCACCCATCTCCACCGCCGGGTTCATCGATCC 360
Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 361 GGGTTCTCCGGGCACGTCAGCTCGAGCTCTCCAACTCCGCCACCTTCGATCAAGCTC 420
Qy 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
Db 421 TGCCCGGGCATGAAGATCGGGCAGCTGTGCATGTTCCGGCTCAGCTCGCCCGCGGATTC 480
Qy 137 ProTyrArgGlyAsnTyrGlnGlySer 145
Db 481 CCCTACGGCAGTACCGCTACGGTTCC 507

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1.86e-13 Length: 9025608
Score: 210.00 Matches: 53
Percent Similarity: 50.30% Conservative: 32
Best Local Similarity: 31.36% Mismatches: 60
Query Match: 26.05% Indels: 24
DB: 15 Gaps: 3

US-08-957-709a-71 (1-156) x US-10-156-761-1 (1-9025608)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluLe-----LeuIleGlu 15
Db 5453355 GTGCTTCTCTCAGACAAGGACATCCGGCCGAGATCGACGCCGAGCGGTTCGATATCGAC 5453414
Qy 16 ProPheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
Db 5453415 CCGTACGACGAATCCATCGGTGTCAGCCCTCGAGCATCGATGTCGGCTCGACCGCTACTTC 5453474
Qy 34 -----GluAlaPheValLys 38
Db 5453475 CGGGTGTTCGAGAATCACCGCTATCCGATATCGACCCCTTCCTCGAGCAGCGCGATCTG 5453534
Qy 39 GlyLysLeuIleAspValGluLysGluGlyLysValIleProProArgGluTyrAla 58
Db 5453535 ACGGCTCTGTCGAGCCCGGAGGCGAGCGCCGCTTCATCTTCATCCGGGGAGTTCGTC 5453594
Qy 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIle 78
Db 5453595 CTCGCCAGTACCTACGAGGTCATCTCCCTCCCGAGCAGCTCGCGTCCGGCTGGAGGGG 5453654
Qy 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
Db 5453655 AAGAGTTCCTCCGGGGCTCGGGCTGGTGGTCACCCATCTCCACCGCCGGGTTCATCGATCC 5453714
Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 5453715 GGGTTCTCCGGGCACGTCAGCTCGAGCTCTCCAACTCCGCCACCTTCGATCAAGCTC 5453774
Qy 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
Db 5453775 TGCCCGGGCATGAAGATCGGGCAGCTGTGCATGTTCCGGCTCAGCTCGCCCGCGGATTC 5453834
Qy 137 ProTyrArgGlyAsnTyrGlnGlySer 145
```


Alignment Scores:

Alignment Scores:		
Pred. No.:	2 069-10	89047
Score:	168.50	49
Percent Similarity:	45.40%	30
Best Local Similarity:	28.16%	Conservative: 60
Query Match:	20.91%	Mismatches: 35
DB:	18	Indels: 35
		Gaps: 6

US-08-957-709A-71 (1-156) x US-10-672-787-34 (1-89047)

Qy 13 LeuIleGluProPheSerCluGluSerIeuGln----- 23
:::||||| ||| :::::
Db 56879 ATGATTAGCCGCTTTGAGCCTGAGCAGGTACGCTAAATGCACAGGTCAAAGTTGGTC 56938

Qy 24 -----ProLaGcTyTrrpLcuArgValGlyArgGlu----- 34
Db 56939 AGCTACGGCACCTCAAGCTATGATGATGATGACGCTGCTATGAATTTAAGTATTT 56998

Qy 35 -----AlaPheValIysGlyLys-----Leulleasp 43
||| ||| ||| |||
Db 56999 ACCAATGTGCATTCAGCCATCGTAGATCCCAAAAATTGTATGATCGCAGATTTTATTGAC 57058

Qy 44 ValGluLysGluGlyLysValValIleProProArgGluTyrAlaLeuLeuLeuThrLeu 63
::
Db 57059 ATCATTGGTGAT---GAATGTATCATTCACCAAAATTCATTGCTTTGGCGGGCACTGTT 57115

Qy 64 GluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeuAla 83
||| ::::||||| |||||::: ::::|||||
Db 57116 GAGTATTTTCGCATCCCAAGAGATGTACTACGACTCTTTGGGCAAAATCCACTTATGCA 57175

Qy 84 ArgGluClyValIleClySerPheAlaTrpValAepProGlyTrrAspGlyAsnLeuThr 103
||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Db 57176 CGCTGTGGAAATCATTTGTCAATGTCACGCCACTTGAGCCCTGAATGGGGGGGCATGTAAACC 57235

Qy 104 LeuMetLeuTy^rAsnAlaSerAsnGluProValGluLeuAryTyr^cGlyGluArgPheVal 123
||||| :||| ::||| ||| |||||
Dβ 57236 TTAGAATTTAGCAATAACACCAAATTGGCAGCACGCATCTATGTGGTGAGGGGTGGCA 57295

Qy 124 GlnIleAlaPheIleArgLeuGluGlyPro-----AlaArgAsnProTyrArg----- 139
|||::: ||| ::: :::
Db 57296 CAGATGCTATTTTTCACATCAGATGCTGATGATGTGCCAAATTTCTTTATAAAGATCGT 57355
||||: ||| ::: :::

Qy 140 ---GlyAsnTyGlnGlySerThrArgLeuAlaPheSerLys 152
||| ||||| ||| :::
Db 57356 GGTGGTAATATCAAGGTCAACAGGTGTACCTTACCAAAA 57397

RESULT 9

Publication No. US20040142413A1

APPLICANT: Reinach, Fernando

INVENTOR: GERS, RICHARD E.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME

FILE REFERENCE: IAFSE 202 US (10213390)
: CURRENT APPLICATION NUMBER: US/10/297,465A

; PRIOR FILING DATE: 2001-06-07
 ; PRIOR AFFILIATION NUMBER: FCI/ID01/CI010

; PRIOR FILING DATE: 2001-06-17
 ; NUMBER OF SEQ ID NOS: 1

```
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465A-1

Alignment Scores:
Pred. No.: 8.38e-08 Length: 2731748
Score: 164.50 Matches: 43
Percent Similarity: 48.23% Conservative: 25
Best Local Similarity: 30.50% Mismatches: 52
Query Match: 20.41% Indels: 21
DB: 19 Gaps: 4

US-08-957-709A-71 (1-156) x US-10-297-465A-1 (1-2731748)
QY 26 GlyTyrAspLeuArgValGlyArgGlu----- 34
Db 716579 GGCTACGAGCGTGGCTGTTCCCGTGAAATTCAGATATTACCCTCAATCAACCACTT 716520
QY 35 -----AlapheVallysglylysLeulleAspValGluGlyGluGlyVal 50
Db 716519 GTCGATCCAAAACAATTCGATATGAAGCTTTCATTTGATGTCGAATCAGAC---GTCGTC 716463
QY 51 ValIleProProArgGluTyrAlaLeulleLeuThrLeuGluArgIleLysLeuProAsp 70
Db 716462 ATCATCCCAACCAAGAGCTTCGCTTGGCAGCCACCATGTAGTATTCGGCATCCACGC 716403
QY 71 AspValMetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySer 90
Db 716402 AATGTACTGGTGATTTGTTCCGAAAGACACCTACGACGTTGTGGAATCATCTCAAT 716343
QY 91 PheAlaTrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSer 110
Db 716342 GTGACCCCAACCAAGAGCTTCGCTTGGCAGCCACCATGTAGTATTCGGCATCCACGC 716283
QY 111 AsnGluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeu 130
Db 716282 CCGTTACCCGACGATCTACCGAAGAGGGGTGTAGCCAGATGTTGTTCTCGAAGCC 716223
QY 131 -----GluGlyProAlaArgAsnProTyrArg-----GlyAsnTyrGlnGlySer 145
Db 716222 GATCCGACGAGCGTGTGCGCAAAATCTCTACCGCAGCGCAACGCGCAAAATATCAAGGACAG 716163
QY 146 Thr 146
Db 716162 ACG 716160

RESULT 10
US-10-784-986-3
; Sequence 3, Application US/10784986
; Publication NO. US20040229311A1
; GENERAL INFORMATION:
; APPLICANT: HIRANO, Seiko
; APPLICANT: YASUEDA, Hisashi
; TITLE OF INVENTION: Novel lysine decarboxylase gene and method for
; FILE REFERENCE: producing L-lysine
; CURRENT APPLICATION NUMBER: US/10/784,986
; PRIOR FILING DATE: 2004-02-25
; PRIOR APPLICATION NUMBER: JP 2003-47185
; PRIOR FILING DATE: 2003-02-25
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2964
; TYPE: DNA
; ORGANISM: Methylophilus methylotrophus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (684)..(2930)
US-10-784-986-3
```

```
Alignment Scores:
Pred. No.: 2.5e-11 Length: 2964
Score: 160.50 Matches: 44
Percent Similarity: 42.11% Conservative: 28
Best Local Similarity: 25.73% Mismatches: 68
Query Match: 19.91% Indels: 31
DB: 20 Gaps: 3

US-08-957-709A-71 (1-156) x US-10-784-986-3 (1-2964)
QY 13 LeulleGluProPheSerGluGluSerLeuGlnProAla----- 25
Db 61 ATGATTGAGCGGTTTGTAGCCCAAGCTTGTACGTGTAGACCAATTTGACCCGACGTCGTTTGTGAGGTC 120
QY 26 -----GlyTyrAspLeuArg----- 30
Db 121 TATGCCACCTCTTCTTACGGTTTACGATATCGTTGTGTGACGAATTCGCGGTATTATTACC 180
QY 31 -----ValGlyArgGluAlaPheValLysGlyLysLeulleAspVal 44
Db 181 AATATCAACAGCACCATAGTTGACCCCAAGCAATTTGACCCGACGTCGTTTGTGAGGTC 240
QY 45 GluLysGluGlyLysValIleProProArgGluTyrAlaLeulleLeuThrLeuGlu 64
Db 241 TCCGCAAAAGGCTATTGCGTGATTCCCTTAACCTATTGCACTGCGCGCACGGTAGAG 300
QY 65 ArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeuAlaArg 84
Db 301 TATTTCCGTATTCCTCGCTGTACTGTACTGTATGCTCGGCAAGTCACTTATGCGCGT 360
QY 85 GluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGlyAsnLeuThrLeu 104
Db 361 TCGCGCATATTGCTCAACGTCACCCCTTTGAACCAAGAGTGGAAAGGCTATGTCACTA 420
QY 105 MetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGluArgPheValGln 124
Db 421 GAGTTCACCAACACCAACCCCTACCCGCAAAATTTATGCTGCGCAAGGCTGTGCGCAA 480
QY 125 IleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArg-----GlyAsn 141
Db 481 GTGCTGTTCTTTGAGTCTGTAGAAATCTGTGAACGAGCTACAAAGACCGTGTGTTGTA 540
QY 142 TyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 541 TACCAGGCTCAAAATTTGGCGTGACCTGCCAAAA 573

RESULT 11
US-10-289-762-1
; Sequence 1, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; FILE REFERENCE: thereof and uses thereof, in particular for the diagnosis, prev
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
```

US-10-289-762-1


```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
```

```
Alignment Scores:
Pred. No.: 8,248-07 Length: 1230025
Score: 154.00 Matches: 46
Percent Similarity: 43.72% Conservative: 34
Best Local Similarity: 25.14% Mismatches: 63
Query Match: 19.11% Indels: 40
DB: 17 Gaps: 6
```

US-08-957-709a-71 (1-156) x US-10-289-762-1 (1-1230025)

```
QY 13 LeuileGluProPhe-----SerGluGluSer----- 21
DB 450349 ATGATCCATCCCTTTGTTAATGCCCAAGTGAACGTAATAGGACAGCGGAAAAAAGCTT 450408
QY 22 -----LeuGlnProAlaGlyTyrAspLeuArgValGlyArgGlu----- 34
DB 450409 ATAAGTTACGGCCTATCGAGTTATGTTACGACCTCCGCTATCTCGAGAATTCAAAGTG 450468
QY 35 -----AlaPheValLysGlyLysLeuile 42
DB 450469 TTCACCAATGCTATAACTCTCTGTTGATCCAAAATGCTTTACTGAGATATCTTCATC 450528
QY 43 AspValGluLysGluGlyLysValValilleProProArgGluTyrAlaLeuileLeuThr 62
DB 450529 TCTATT---ACTGATGACGCTGTATTGTTCTCCAAAATTCCTTTCCTAGCTCGTAGC 450585
QY 63 LeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeu 82
DB 450586 GTTGAGTATTTCGGAATTCCTAGAAAATGCTTTAACAAATGTGTATAGGAAGTCTACATAT 450645
QY 83 AlaArgGluGlyValilleGlySerPheAlaTrpValAspProGlyTyrAspGlyValLeu 102
DB 450646 GCACGCTGTGGAATTCGTAATATGTCACACCTTTTGGACCTCGAATGGGAAGGCGATGTG 450705
QY 103 ThrLeuMetLysAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlyArgPhe 122
DB 450706 ACTATAGAAATTCCTAACACTAGCCATTCGCCAGCGAAAATTCAGCTAATGAAGGGATT 450765
QY 123 ValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyr-----Arg 139
DB 450766 GCCCAGGCTTATTCTTTGAGTCTAGTACGACCTGCGAGGTTTCTTATGTCAGACAGAAA 450825
QY 140 GlyAsnTyrGlnGlySerThrArgLeu-----AlaPheSerLysArg 153
```

```
Db 450826 GGAAAGTATCAAAAGCAACAGGCATCACCGTACTCTGTGTCTAAAGTTTCAGTAAGAAA 450885
QY 154 LysLysLeu 156
DB 450886 AAAAAAATCG 450894
RESULT 12
US-10-437-963-14349
; Sequence 14349, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 14349
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_20296C.1
US-10-437-963-14349
```

```
Alignment Scores:
Pred. No.: 9,29e-11 Length: 586
Score: 149.50 Matches: 42
Percent Similarity: 50.00% Conservative: 18
Best Local Similarity: 35.00% Mismatches: 56
Query Match: 18.55% Indels: 4
DB: 19 Gaps: 2
```

US-08-957-709a-71 (1-156) x US-10-437-963-14349 (1-586)

```
QY 22 LeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLysLeu 41
DB 160 TTAAGTGGCCTGGTGAACAAGTCAGCCACAACTTGAAAAAATCATGATGATGAAT 219
QY 42 IleAspValGluLysGluGlyLysValValilleProProArgGluTyrAlaLeuileLeu 61
DB 220 ATTTAGCAGAGGTGAAGCT---TTTTCTTCATCCAGCGGAATTAGCCTTGGCCACT 276
QY 62 ThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSer 81
DB 277 ACCCTAGAGTCAGTGAATAATACCGGATAATATTGTTGGTGGCTAGATGAACGTTCTTCA 336
QY 82 LeuAlaArgGluGlyValilleGlySerPheAla-----TrpValAspPro-GlyTrpAs 99
DB 337 TTGGCTCCCTTAGGATTAATGATCATGTTACCGCTCATCGCATTTGATCCAGGTTGGCA 396
QY 99 pGlyAsnLeuThrLeuMetLysAsnAlaSerAsnGluProValGluLeuArgTyrGln 119
DB 397 CGGCAAAATTTGATTAGAAATTCCTTAATGCTGGTAAACTACCATTTAGCCTTCGCTTAA 456
QY 119 yGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyr 138
DB 457 TATGCGGATGTTGATTAAGCTTTGAAGTATTGAGTGGAGCAGCGGCTTAAGCCCTTAT 514
RESULT 13
US-10-329-670-1
; Sequence 1, Application US/10329670
; Publication No. US20040018503A1
; GENERAL INFORMATION:
```

APPLICANT: Fleischmann et al.
 TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
 FILE REFERENCE: PB186P1
 CURRENT APPLICATION NUMBER: US/10/329,670
 CURRENT FILING DATE: 2002-12-24
 PRIOR APPLICATION NUMBER: US 09/643,990
 PRIOR FILING DATE: 2000-08-23
 PRIOR APPLICATION NUMBER: US 08/487,429
 PRIOR FILING DATE: 1995-06-07
 PRIOR APPLICATION NUMBER: US 08/426,787
 PRIOR FILING DATE: 1995-04-21
 NUMBER OF SEQ ID NOS: 1
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 1
 LENGTH: 1830121
 TYPE: DNA
 ORGANISM: Haemophilus influenzae
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (4747)..(4747)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (9921)..(9921)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (10150)..(10150)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (29298)..(29298)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (36543)..(36543)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (36636)..(36636)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (40808)..(40810)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (44416)..(44416)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (44905)..(44905)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (44975)..(44975)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (45593)..(45593)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (45732)..(45732)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (47036)..(47036)

```
;
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152530)..(152530)
```

```
Alignment Scores:
Pred. No.: 6,07e-06 Length: 1830121
Score: 149.50 Matches: 44
Percent Similarity: 45.93% Conservative: 24
Best Local Similarity: 29.73% Mismatches: 53
Query Match: 18.55% Indels: 27
DB: 17 Gaps: 5
```

US-08-957-709A-71 (1-156) x US-10-329-670-1 (1-1830121)

```
QY 18 SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
   :::::
Db 146725 AATAATGACAAAATCAACGGTCGCAATGTGCTGGGGAATCTTTCGGTGTA 146784
   :::::
QY 38 -----LySGlyLysLeuLeuAspValGlu-----LysGlu----- 47
   :::::
Db 146785 TTTCGTGAACACTCTGCACCTTTTATTGATCTCAAGCGGCCGAAAGAAGATATCCGCC 146844
   :::::
QY 48 -----GlyLysValValIlePro----- 53
   :::::
Db 146845 CAGCTTGATCATGTAATGAGCATGAAATATATTCAGAGGGGGAGACATTTTCTTA 146904
   :::::
```

```
QY 54 ---ProArgGluTyrAlaLeuLeuLeuThrLeuGluArgIleLysLeuProAspAspVal 72
   :::::
Db 146905 CATCTGTGTAATTTAGCTTAGCAACACCGCTTGAATCCGTGAATTTGCCAGCAATATT 146964
   :::::
QY 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
   :::::
Db 146965 ATCGTTGGTTAGATGCGCTTCTTTAGCCCGTTTAGCTTAATGGTACACGTACACC 147024
   :::::
QY 93 -----TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSer 110
   :::::
Db 147025 GCACACCGTATTGATCCAGGTGGGAAGAAATCGTCTGGAATTTTAAATTCAGGG 147084
   :::::
QY 111 AsnGluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeu 130
   :::::
Db 147085 AAATTACATTTAGCATTTACGCCCAATATGTAATTTGGTGGTTAAGTTTTGAAGTATTA 147144
   :::::
QY 131 GluGlyProAlaArgAsnProTyr 138
   :::::
Db 147145 AGTGGCGAASSAAAAACGACCATAC 147168
   :::::
RESULT 14
US-10-158-865-1
; Sequence 1, Application US/10158865
; Publication No. US20040203093A1
;
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P2C1D1
; CURRENT APPLICATION NUMBER: US/10/158,865
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
```

NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature

LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)

```
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
```

```
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.: 6.07e-06 Length: 1830121
Score: 149.50 Matches: 44
Percent Similarity: 45.95% Conservative: 24
Best Local Similarity: 29.73% Mismatches: 53
Query Match: 18.55% Indels: 27
DB: 20 Gaps: 5

US-08-957-709A-71 (1-156) x US-10-158-865-1 (1-1830121)

Qy 18 SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
Db 146725 AATAATGACAAAATCAACGGTCCGACAAATGATGTTCTGCTGGGGAATTTCTTCGTGA 146784
Qy 38 -----LysGlyLysLeuIleAspValGlu-----LysGlu----- 47
Db 146785 TTTCGTGAACACTCTGCACCTTTTATTGATCTAAGCGGCCGCGAAAGAAAGTATCCGCC 146844
Qy 48 -----GlyLysValValIlePro----- 53
Db 146845 CAGCTTGAATCNGTAATGAGCGATGAATATTATTCAGAGGGGGAAGCATTTTCTTCA 146904
Qy 54 ---ProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspAspVal 72
Db 146905 CATCTGGTACTTTAGCCCTTAGCACCAACCGTTGAATCCGTGAATTCGCGAATTCGACCAATATT 146964
Qy 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 146965 ATCGGTTGGTAGATGGCGGTTCTTCTTTAGCCGCTTATAGGCTTAATGGTACACGTCACC 147024
Qy 93 -----TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSer 110
Db 147025 GCACACCGTATTGATCCAGGTGGGAAGGAAAAATCGTGCTGGAATTTATAATTCAGGG 147084
Qy 111 AsnGluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeu 130
Db 147085 AAATTACCATTTAGCATACGCCCAATATGGTAATTTGGTGGCTTAAGTTTGAAGTATTA 147144
Qy 131 GluGlyProAlaArgAsnProTyr 138
Db 147145 AGTGGCGAASSAAAAACGACCATAC 147168

RESULT 15
US-10-981-687-1
; Sequence 1, Application US/10981687
; Publication No. US20050131222A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P2C1D12
; CURRENT APPLICATION NUMBER: US/10/981,687
; CURRENT FILING DATE: 2004-11-05
; PRIOR APPLICATION NUMBER: US 10/158,865
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
```



```
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
```

```
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:

Alignment Scores:
Pred. No.: 6.07e-06 Length: 1830121
Score: 149.50 Matches: 44
Percent Similarity: 45.95% Conservative: 24
Best Local Similarity: 29.73% Mismatches: 53
Query Match: 18.55% Indels: 27
DB: 22 Gaps: 5

US-08-957-709A-71 (1-156) x US-10-981-687-1 (1-1830121)

QY 18 SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
Db 146725 AATAATGACAAATCAACGGTCGACAATTGATGTTCTGGGGAATTCTTTCCGTGTA 146784
QY 38 -----LysGlyLysLeuIleAspValGlu-----LysGlu----- 47
Db 146785 TTTCGTGAACACTCTGCACCTTTTATTGATCTAAGCGCGCCGAAAGAGATATCGCC 146844
QY 48 -----GlyLysValIlePro----- 53
Db 146845 CAGCTTGAATCAGTAATAGCGATGAAATTATTATTCAGAGGGGGAAGCATTTTCTTA 146904
QY 54 ---ProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 146905 CATCTGTGTACTTTAGCCCTTAGCAACACCGTGAATCCGTGAAATTCGACCAATATT 146964
QY 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 146965 ATCGGTGGTTAGATGGCGTCTCTTTTAGCCCGTTTAGCTTAATGTCACGTCACC 147024
QY 93 -----TrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSer 110
Db 147025 GCACACCGTATTGATCCAGTTGGGAAGGAAAAATCGTCTGGAATTTTATAATTCAGG 147084
QY 111 AsnGluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeu 130
Db 147085 AAATTACCATTAGCATTCAGCCCAAAATATGTAATTGGTCGTTAGTTTGAAGTATTA 147144
QY 131 GluGlyProAlaArgAsnProTyr 138
Db 147145 AGTGGGGAASSAAAAAGCACCATAC 147168
```

Search completed: July 2, 2005, 13:18:17
Job time : 3989 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 2, 2005, 08:11:01 ; Search time 473 Seconds
(without alignments)
1952.388 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPWKIRKEILIEPFSEE.....PYRGNVQSGSTRLAFSKRKL 156

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool_08957709/runat_01072005_154631_25999/app_query.fasta_1.327
-DB=N Geneseq 16Dec04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOFCU=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US08957709 @CNC 1 1 644 @runat_01072005_154631_25999 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004as:*
13: Geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	806	100.0	471	2 AAV63860	AAV63860 Polymers
2	703	87.2	265118	5 AAH41227	AAH41227 Pyrococcus
c 3	696	86.4	110000	12 ADN46845_12	Continuation (13 o
c 4	696	86.4	110000	12 ADN47591_08	Continuation (9 of
c 5	696	86.4	110000	12 ADN46123_12	Continuation (13 o

c 6	696	86.4	110000	12 ADN47209_08	Continuation (9 of
c 7	696	86.4	110000	12 ADN46464_12	Continuation (13 o
c 8	696	86.4	110000	12 ADN47960_08	Continuation (9 of
c 9	242	30.0	3382	2 AAQ62300	AAQ62300 p2am-l3 p
10	227.5	28.2	597	12 ADK16664	Adk16664 Nanoarcha
c 11	227.5	28.2	110000	12 ADK16049_2	Continuation (3 of
c 12	220	27.3	110000	11 ADM27081_02	Continuation (3 of
c 13	214	26.6	567	5 AAH68105	AAH68105 C glutami
c 14	214	26.6	309400	5 AAH68534	AAH68534 C glutami
c 15	214	26.6	349980	5 AAH68533	AAH68533 C glutami
c 16	196	24.3	580	4 AAF72258	AAf72258 Corynebac
c 17	194	24.1	423	6 ABK77720	ABk77720 Bacillus
c 18	194	24.1	110000	4 AAI99682_03	Continuation (4 of
c 19	194	24.1	110000	4 AAI99683_03	Continuation (4 of
c 20	191.5	23.8	349980	6 ABQ81849	ABq81849 Bifidobac
c 21	185.5	23.0	594	9 ADA29252	ADA29252 DNA encod
c 22	181	22.5	2006	2 AAV90555	AAv90555 Nucleotid
c 23	181	22.5	2588	2 AAV90871	AAv90871 Nucleotid
c 24	174	21.6	110000	6 ABA92787_1	Continuation (2 of
c 25	172	21.3	609	11 ABD02487	ABd02487 Pseudomon
c 26	170.5	21.2	582	10 ACF68843	ACf68843 Photorhab
c 27	170.5	21.2	110000	10 ACF67367_16	Continuation (17 o
c 28	170.5	21.2	249878	10 ACF65381	ACf65381 Photorhab
c 29	168.5	20.9	579	12 ADL04075	Adl04075 DNA encod
c 30	168.5	20.9	615	10 ADF00606	Adf00606 Bacterial
c 31	168.5	20.9	89047	4 AAF28547	AAf28547 Genomic f
c 32	164	20.3	654	11 ACH95333	ACH95333 Klebsiell
c 33	164	20.3	110000	11 ADM27081_15	Continuation (16 o
c 34	160.5	19.9	2964	13 ADR89139	Adr89139 Nucleotid
c 35	156	19.4	564	10 ABZ39371	ABz39371 N. gonorr
c 36	155.5	19.3	419	6 ABX66923	ABx66923 Helicobac
c 37	155	19.2	78845	3 AAA81463	AAA81463 N. mening
c 38	155	19.2	110000	3 AAA81490_08	Continuation (9 of
c 39	155	19.2	349980	3 AAF21608	AAf21608 Neisseria
c 40	154.5	19.2	573	12 ADQ57545	AdQ57545 Actinobac
c 41	154	19.1	110000	2 AAV21209_10	Continuation (11 o
c 42	154	19.1	110000	2 AAX91390_04	Continuation (5 of
c 43	154	19.1	273254	3 AAC81314	AAc81314 Chlamydia
c 44	152.5	18.9	4037	13 ADT05374	Adt05374 Haemophil
c 45	152.5	18.9	349980	13 ADT05648	Adt05648 Haemophil

ALIGNMENTS

RESULT 1
AAV63860
ID AAV63860 standard; DNA; 471 BP.
XX
AC AAV63860;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 (dUTPase) component DNA.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication; ss.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
(STRA-) STRATAGENE.
PI Hogrefe H, Hansen CJ;
XX

DR WPI: 1998-542284/46.
DR P-PSDB; AAW72847.
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX
XX
PS Claim 17; Page 43; 161pp; English.
XX
CC This DNA sequence encodes the P45 component (see AAW72847) of the
CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. PCR
CC primers (see AAW63861-64) based on an N-terminal peptide (see AAW72846)
CC of P45 were used to amplify P. furiosus genomic DNA, and sequencing was
CC carried out on purified PCR products and plasmid mini-preps. P45 and P40
CC (see AAW72844) are the predominant components of PEF, which acts to
CC enhance the activity of P. furiosus DNA polymerase, thereby providing
CC replication products of greater length and purity. P45 was identified as
CC a dUTPase, and possesses polymerase enhancing activity. The invention
CC provides novel extracts, proteins and complexes that improve the
CC polymerisation activity of nucleic acid polymerases, as well as DNA
CC constructs and antibodies. Also included are methods for identifying
CC compositions with polymerase enhancing activity, for purifying and using
CC these compositions, and specific extracts, proteins and complexes that
CC function to enhance polymerase activity. Nucleic acid polymerase
CC reactions can be enhanced (claimed) by mixing a nucleic acid template, at
CC least 1 polymerase and a composition having polymerase enhancing
CC activity. Kits are provided for replicating nucleic acids. The kits can
CC be used in site-directed mutagenesis, nucleic acid sequencing or
CC amplification (preferably PCR or RT-PCR). Isolated DNA can be used in the
CC recombinant production of P45. (Updated on 17-OCT-2003 to standardise OS
CC field)
XX
XX Sequence 471 BP; 156 A; 80 C; 124 G; 111 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.7e-97 Length: 471
Score: 806.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-08-957-709A-71 (1-156) x AAW63860 (1-471)

QY 1 MetLeuLeuProAspTrrpLysLeuArgLysGluLeuLeuGluProPheSerGluGlu 20
DB 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACATTATAGAGCCATTTCAGAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAAACGACGAGGTATACCTCAGAGTGGCGAGAGAGGCTTTTGTAAAGGGGAAA 120
QY 41 LeuileAspValGluLysGluGlyLysValValileProProArgGluTyrAlaLeuile 60
DB 121 TTAATCGACGCTGGAAAGGAAAGGAAAGTCGTTATTCCTCCAAAGGGAATACGCCCTAATC 180
QY 61 LeuThrLeuGluArgLileLysLeuProAspValMetGlyAspMetLysileArgSer 80
DB 181 CTAACCCCTCGAGAGGATAAAGTGGCCGACGATGTTATGGGGGATATGAAGATAGGAGC 240
QY 81 SerLeuAlaArgGluGlyValileGlySerPheAlaTrrpValAspProGlyTrrpAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTTATGGTCTTTTGTCTGGGTTGACCCAGGATGGATGGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACACTAATGCTCTACAATGCCTCAATGAACCTGTGCAATTAAGATATGAGAG 360
QY 121 ArgPheValGlnIleAlaPheileArgLeuGluGlyProAlaArgAsnProTrrpArgGly 140
DB 361 AGATTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLeu 156

DB 421 AACTATCAGGGGAGCACAAAGGTAGCGTTTTCAGAGAGAAAGAACTC 468

RESULT 2

AAH41227
ID AAH41227 standard; DNA; 265118 BP.
XX AC AAH41227;
XX DT 29-OCT-2001 (first entry)
XX DE Pyrococcus abyssi genomic fragment #6.
XX KW Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX OS Pyrococcus abyssi.
XX FH Key Location/Qualifiers
FT misc_feature 1..49980
FT /tag= a
FT /note= "This sequence overlaps with the 3' end of
FT AAH41226"
XX FR2792651-Al.
XX PD 27-OCT-2000.
XX PF 21-APR-1999; 99FR-00005034.
XX PR 21-APR-1999; 99FR-00005034.
XX PA (CNRS) CNRS CENT NAT RECH SCI.
XX PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins
XX useful in industry.
XX Claim 1; Page 593-665; 1657pp; French.

CC The present invention relates to the genomic sequence of *Pyrococcus*
CC abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC vents. The present sequence is a fragment of the genomic sequence of P.
CC abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41226.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade. Note: This patent is in the same patent family as
CC WO2000085082, which contains additional sequences as shown in AAB99132-
CC AAB99143, AAH75903-AAH75920 and AAG66436
XX SQ Sequence 265118 BP; 75226 A; 61302 C; 54005 G; 74585 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.36e-80 Length: 265118
Score: 703.00 Matches: 131
Percent Similarity: 94.12% Conservative: 13
Best Local Similarity: 85.62% Mismatches: 9
Query Match: 87.22% Indels: 0
DB: 5 Gaps: 0

US-08-957-709A-71 (1-156) x AAH41227 (1-265118)

QY 1 MetLeuLeuProAspTrrpLysLeuArgLysGluLeuLeuGluProPheSerGluGlu 20
DB 262045 ATGCTCTCTCCAGACTGGAAATAGGAAGAGATTTTAAAGCAATTCACAGAGAA 262104
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40

Db 262105 TCGCTTCAACAGCAGCTGGCTACGACTTAAGGGTAGGCAAGGAAGCTTATCATCTTCAGGGAAAG 262164

Qy 41 LeuileaspValGluLysGluGlyLeuValPheProArgGluTyrAlaLeuile 60
|||||
262165 TTCATAGATTGTGAAGAAGGAGGGCAAAGTCATAATACCTCCAAAAGAATAATGCCCTAATA 262224

Qy 61 LeuThrLeuGluArGlleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
|||||
262225 CTGACCCTTGAGAGGATAAAGCTTCCAGATGCATATTATGGGAGACATGAAGATAAGGAGC 262284

Qy 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100
|||||
262285 AGCTAGCTAGGAGAGGTGTTCTAGGCTCTCTTGCATGGGTAGACCCGGGATGGGACCGT 262344

Qy 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
|||||
262345 AATCTAACCTTGATGCTTTACAATGCATCGAAAAAGGAGGTAAATTTTAAGGTACAAAGAG 262404

Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
|||||
262405 AGGTTTTCTCAGATAGGCTTCTTAAGGCTTGAAGCTCCCGCCAAGAAATCCATACATGAGGGC 262464

Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArg 153
|||||
262465 AACATCAAGGAAGCAGAGAAATAGTCTCTCAAAAGAGA 262503

RESULT 3
ADN46845_12/C
Continuation (13 of 21) of ADN46845 from base 1200001 (Thermococcus kodakaraensis KOD1)
WP Sequence split into 21 fragments LOCUS ADN46845 Accession Adn46845

Fragment Name	Begin	End
WP ADN46845_00	1	110000
WP ADN46845_01	100001	210000
WP ADN46845_02	200001	310000
WP ADN46845_03	300001	410000
WP ADN46845_04	400001	510000
WP ADN46845_05	500001	610000
WP ADN46845_06	600001	710000
WP ADN46845_07	700001	810000
WP ADN46845_08	800001	910000
WP ADN46845_09	900001	1010000
WP ADN46845_10	1000001	1100000
WP ADN46845_11	1100001	1210000
WP ADN46845_12	1200001	1310000
WP ADN46845_13	1300001	1410000
WP ADN46845_14	1400001	1510000
WP ADN46845_15	1500001	1610000
WP ADN46845_16	1600001	1710000
WP ADN46845_17	1700001	1810000
WP ADN46845_18	1800001	1910000
WP ADN46845_19	1900001	2010000
WP ADN46845_20	2000001	2089378

Alignment Scores:
Pred. No.: 1.58e-79 Length: 110000
Score: 696.00 Matches: 130
Percent Similarity: 92.26% Conservative: 13
Best Local Similarity: 83.87% Mismatches: 12
Query Match: 86.35% Indels: 0
DB: 12 Gaps: 0

US-08-957-709A-71 (1-156) x ADN46845_12 (1-110000)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuileGluProPheSerGluGlu 20
|||||

Db 40800 ATGATGCTCCCCGATTGGAAGATTAGAAAGAAATTTTGATCGAGCCTTTTCAGTGAAGAA 40741

Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
|||||

Db 40740 TCCCTGCGAGCTGCGAGATATGACCTAAGGGTTGGAGCGAGGCTTACGTAATGGTAAA 40681

Qy 41 LeuileaspValGluLysGluGlyLysValIleProArgGluTyrAlaLeuile 60
|||||

Db 40680 ATACTCGAGTTAAGGATTACAGGGGAGTTACAAATACCCCAAGACTTACGCCTTGGTT 40627
 Qy 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
 Db 40620 TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAAATGGGGGACATGAAGCTTAGGAGC 40561
 Qy 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTtpValAspProGlyTtpAspGly 100
 Db 40560 AGTCTAGCCAGAGGGCTTAATCGGTTCTTTTGGCTGGGTTCAGCCCTGGATGGATGGA 40501
 Qy 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
 Db 40500 AATCTCACTCTAGCCCTGTTCACAGCTTCACAGAGTCTGTGGAACTCAATACGGCGAG 40441
 Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaIleArgAsnProTyrArgGly 140
 Db 40440 CGCTTTGTTTCAGATACGCTTTATCGGCTGGAGGGGCGCTGCTAAAAACCCCTATCGTGA 40381
 Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys 155
 Db 40380 AATTACAGGGAAGCAACATCTAGCGCTCTCAAAAGGAAACGG 40336

RESULT 4
 ADN47591_08
 Continuation (9 of 21) of ADN47591 from base 800001 (Thermococcus kodakaraensis
 WP Sequence split into 21 fragments LOCUS ADN47591 Accession Adn47591
 WP Fragment Name Begin End
 WP ADN47591_00 1 110000
 WP ADN47591_01 100001 210000
 WP ADN47591_02 200001 310000
 WP ADN47591_03 300001 410000
 WP ADN47591_04 400001 510000
 WP ADN47591_05 500001 610000
 WP ADN47591_06 600001 710000
 WP ADN47591_07 700001 810000
 WP ADN47591_08 800001 910000
 WP ADN47591_09 900001 1010000
 WP ADN47591_10 1000001 1110000
 WP ADN47591_11 1100001 1210000
 WP ADN47591_12 1200001 1310000
 WP ADN47591_13 1300001 1410000
 WP ADN47591_14 1400001 1510000
 WP ADN47591_15 1500001 1610000
 WP ADN47591_16 1600001 1710000
 WP ADN47591_17 1700001 1810000
 WP ADN47591_18 1800001 1910000
 WP ADN47591_19 1900001 2010000
 WP ADN47591_20 2000001 2089378

Alignment Scores:
 Pred. No.: 1,58e-79 Length: 110000
 Score: 696.00 Matches: 130
 Percent Similarity: 92.26% Conservatives: 13
 Best Local Similarity: 83.87% Mismatches: 12
 Query Match: 86.35% Indels: 0
 DB: 12 Gaps: 0

US-08-957-709A-71 (1-156) x ADN47591_08 (1-110000)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu 20
 Db 48578 ATGATGCTCCCGATTGGAAGATTAGAAAGGAAATTTTGTATCGAGCCTTTCAGTAGAANA 48637
 Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
 Db 48638 TCCTTCGAGCTTCAGGATATGACCTAAGGGTTGGAGGGAGAGCTTACGTAATCGTANA 48697
 Qy 41 LeuIleAspValGluLysGluGlyLysValIleProProArgGluTyrAlaLeuIle 60
 Db 48698 ATACTCGAGTTAAGGATTCAGGGGAGTTTACAAATACCCCAAGACTTACGCCTTGGTT 48757
 Qy 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80

Db	48758	TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAATAGGGCGACATGAAGCTTAGGAGC	48817
Qy	81	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly	100
Db	48818	AGTCTAGCCAGAGAGGCTTAATCGGTTCTTTTGGCTGGGTTCAGCCCTGGATGGATGA	48877
Qy	101	AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120
Db	48878	AATCTCACTCTAGCCCTGTTCAACGCTTCCACAGAGTCTGTGGAACTCAAATACGGCGAG	48937
Qy	121	ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly	140
Db	48938	CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCCCTGCTAAAAACCCCTATCGTGA	48997
Qy	141	AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys	155
Db	48998	AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAGGAAACGG	49042
RESULT 5			
ADN46123_12/c			
Continuation (13 of 21) of ADN46123 from base 1200001 (Thermococcus kodakaraensis KOD1 9			
WP Sequence split into 21 fragments LOCUS ADN46123 Accession Adn46123			
WP	Fragment Name	Begin	End
WP	ADN46123_00	1	110000
WP	ADN46123_01	100001	210000
WP	ADN46123_02	200001	310000
WP	ADN46123_03	300001	410000
WP	ADN46123_04	400001	510000
WP	ADN46123_05	500001	610000
WP	ADN46123_06	600001	710000
WP	ADN46123_07	700001	810000
WP	ADN46123_08	800001	910000
WP	ADN46123_09	900001	1010000
WP	ADN46123_10	1000001	1110000
WP	ADN46123_11	1100001	1210000
WP	ADN46123_12	1200001	1310000
WP	ADN46123_13	1300001	1410000
WP	ADN46123_14	1400001	1510000
WP	ADN46123_15	1500001	1610000
WP	ADN46123_16	1600001	1710000
WP	ADN46123_17	1700001	1810000
WP	ADN46123_18	1800001	1910000
WP	ADN46123_19	1900001	2010000
WP	ADN46123_20	2000001	2089378
Alignment Scores:			
Pred. No.:	1.58e-79	Length:	110000
Score:	696.00	Matches:	130
Percent Similarity:	92.26%	Conservative:	13
Best Local Similarity:	83.87%	Mismatches:	12
Query Match:	86.35%	Indels:	0
DB:	12	Gaps:	0
US-08-957-709a-71 (1-156) x ADN46123_12 (1-110000)			
Qy	1	MetLeuLeuProAspTrpLysIleArgLysGluLeuLeuLeuGluProPheSerGluGlu	20
Db	40800	ATGATGCTTCCCGATTGGAAGATTAGAAAGGAAATTTTTCAGTCAGCCCTTCAGTGAAGAA	40741
Qy	21	SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys	40
Db	40740	TCCTGCGACCTTCAGGATATGACCTAAGGGTTGGAGGGCGAAGCTTACGTAATGGTAAA	40681
Qy	41	LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle	60
Db	40680	ATACTCGAGCTTAAGGATTTCAGGGGAGTTTACAATACCCCCAAGACTTACGCCCTGGTT	40621
Qy	61	LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleArgSer	80
Db	40620	TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAATGGGCGACATGAAGCTTAGGAGC	40561
Qy	81	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly	100

Db	40560	AGTCTAGCCAGAGAGGCTTAATCGGTTCTTTTGGCTGGGTTCAGCCCTGGATGGATGA	40501
Qy	101	AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120
Db	40500	AATCTCACTCTAGCCCTGTTCAACGCTTCCACAGAGTCTGTGGAACTCAAATACGGCGAG	40441
Qy	121	ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly	140
Db	40440	CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCCCTGCTAAAAACCCCTATCGTGA	40381
Qy	141	AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys	155
Db	40380	AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAGGAAACGG	40336
RESULT 6			
ADN47209_08			
Continuation (9 of 21) of ADN47209 from base 800001 (Thermococcus kodakaraensis KOD1 9er			
WP Sequence split into 21 fragments LOCUS ADN47209 Accession Adn47209			
WP	Fragment Name	Begin	End
WP	ADN47209_00	1	110000
WP	ADN47209_01	100001	210000
WP	ADN47209_02	200001	310000
WP	ADN47209_03	300001	410000
WP	ADN47209_04	400001	510000
WP	ADN47209_05	500001	610000
WP	ADN47209_06	600001	710000
WP	ADN47209_07	700001	810000
WP	ADN47209_08	800001	910000
WP	ADN47209_09	900001	1010000
WP	ADN47209_10	1000001	1110000
WP	ADN47209_11	1100001	1210000
WP	ADN47209_12	1200001	1310000
WP	ADN47209_13	1300001	1410000
WP	ADN47209_14	1400001	1510000
WP	ADN47209_15	1500001	1610000
WP	ADN47209_16	1600001	1710000
WP	ADN47209_17	1700001	1810000
WP	ADN47209_18	1800001	1910000
WP	ADN47209_19	1900001	2010000
WP	ADN47209_20	2000001	2089378
Alignment Scores:			
Pred. No.:	1.58e-79	Length:	110000
Score:	696.00	Matches:	130
Percent Similarity:	92.26%	Conservative:	13
Best Local Similarity:	83.87%	Mismatches:	12
Query Match:	86.35%	Indels:	0
DB:	12	Gaps:	0
US-08-957-709a-71 (1-156) x ADN47209_08 (1-110000)			
Qy	1	MetLeuLeuProAspTrpLysIleArgLysGluLeuLeuLeuGluProPheSerGluGlu	20
Db	48578	ATGATGCTTCCCGATTGGAAGATTAGAAAGGAAATTTTTCAGTCAGCCCTTCAGTGAAGAA	48637
Qy	21	SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys	40
Db	48638	TCCTGCGACCTTCAGGATATGACCTAAGGGTTGGAGGGCGAAGCTTACGTAATGGTAAA	48697
Qy	41	LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle	60
Db	48698	ATACTCGAGCTTAAGGATTTCAGGGGAGTTTACAATACCCCCAAGACTTACGCCCTGGTT	48757
Qy	61	LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleArgSer	80
Db	48758	TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAATGGGCGACATGAAGCTTAGGAGC	48817
Qy	81	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly	100
Db	48818	AGTCTAGCCAGAGAGGCTTAATCGGTTCTTTTGGCTGGGTTCAGCCCTGGATGGATGA	48877
Qy	101	AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120

Db 48878 AATCTCACTAGCCCTGTTCAAGCTTCCACGAGTCTGTGGAACTCAATACGGCGAG 48937

Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140

Db 48938 CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCGCTGCTAAACACCCCTATCGTGA 48997

Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys 155

Db 48998 AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAGGAACCG 49042

RESULT 7

ADN46464_12/c
Continuation (13 of 21) of ADN46464 from base 1200001 (Thermococcus kodakaraensis KOD1
WP Sequence split into 21 fragments LOCUS ADN46464 Accession ADN46464

Fragment Name	Begin	End
WP ADN46464_00	1	110000
WP ADN46464_01	100001	210000
WP ADN46464_02	200001	310000
WP ADN46464_03	300001	410000
WP ADN46464_04	400001	510000
WP ADN46464_05	500001	610000
WP ADN46464_06	600001	710000
WP ADN46464_07	700001	810000
WP ADN46464_08	800001	910000
WP ADN46464_09	900001	1010000
WP ADN46464_10	1000001	1100000
WP ADN46464_11	1100001	1210000
WP ADN46464_12	1200001	1310000
WP ADN46464_13	1300001	1410000
WP ADN46464_14	1400001	1510000
WP ADN46464_15	1500001	1610000
WP ADN46464_16	1600001	1710000
WP ADN46464_17	1700001	1810000
WP ADN46464_18	1800001	1910000
WP ADN46464_19	1900001	2010000
WP ADN46464_20	2000001	2089378

Alignment Scores:
Pred. No.: 1.58e-79 Length: 110000
Score: 696.00 Matches: 130
Percent Similarity: 92.26% Conservative: 13
Best Local Similarity: 83.87% Mismatches: 12
Query Match: 86.35% Indels: 0
DB: 12 Gaps: 0

US-08-957-709a-71 (1-156) x ADN46464_12 (1-110000)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu 20

Db 40800 ATGATGCTTCCCGATTGGAGATTAGAAAGGAATTTTGATCGAGCCTTTCAGTGAAGAA 40741

Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40

Db 40740 TCCTCGAGCCTCGAGATATGACCTTAGGGTTGGAGGCGAAGCTTACGTAATGGTAA 40681

Qy 41 LeuIleAspValGluLysGluGlyLysValIleProProArgGluTyrAlaLeuIle 60

Db 40680 ATACTCGAGTTAAGGATTCCAGGGGAGTTTACAATACCCCAAGACTTACGCCCTGGTT 40621

Qy 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80

Db 40620 TTGACTCTCGAGAGGATCAAGCTTCCGGATGATGTAATGGCGACATGAAGCTTAGGAGC 40561

Qy 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100

Db 40560 AGTCTAGCCAGAGAGGCGCTTAATCGGTTCTTTTGGCTGGGTTGACCTGGATGGGATGA 40501

Qy 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120

Db 40500 AATCTCACTAGCCCTGTTCAAGCTTCCACGAGTCTGTGGACTCAAAATACGGCGAG 40441

Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140

Db 40440 CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCGCTGCTAAACACCCCTATCGTGA 40381

Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys 155

Db 40380 AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAGGAACCG 40336

RESULT 8

ADN47960_08
Continuation (9 of 21) of ADN47960 from base 800001 (Thermococcus kodakaraensis KOD1 ge
WP Sequence split into 21 fragments LOCUS ADN47960 Accession ADN47960

Fragment Name	Begin	End
WP ADN47960_00	1	110000
WP ADN47960_01	100001	210000
WP ADN47960_02	200001	310000
WP ADN47960_03	300001	410000
WP ADN47960_04	400001	510000
WP ADN47960_05	500001	610000
WP ADN47960_06	600001	710000
WP ADN47960_07	700001	810000
WP ADN47960_08	800001	910000
WP ADN47960_09	900001	1010000
WP ADN47960_10	1000001	1110000
WP ADN47960_11	1100001	1210000
WP ADN47960_12	1200001	1310000
WP ADN47960_13	1300001	1410000
WP ADN47960_14	1400001	1510000
WP ADN47960_15	1500001	1610000
WP ADN47960_16	1600001	1710000
WP ADN47960_17	1700001	1810000
WP ADN47960_18	1800001	1910000
WP ADN47960_19	1900001	2010000
WP ADN47960_20	2000001	2089378

Alignment Scores:
Pred. No.: 1.58e-79 Length: 110000
Score: 696.00 Matches: 130
Percent Similarity: 92.26% Conservative: 13
Best Local Similarity: 83.87% Mismatches: 12
Query Match: 86.35% Indels: 0
DB: 12 Gaps: 0

US-08-957-709a-71 (1-156) x ADN47960_08 (1-110000)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu 20

Db 48578 ATGATGCTTCCCGATTGGAGATTAGAAAGGAATTTTGATCGAGCCTTTCAGTGAAGAA 48637

Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40

Db 48638 TCCTCGAGCCTCGAGATATGACCTTAGGGTTGGAGGCGAAGCTTACGTAATGGTAA 48697

Qy 41 LeuIleAspValGluLysGluGlyLysValIleProProArgGluTyrAlaLeuIle 60

Db 48698 ATACTCGAGTTAAGGATTTCAGGGGAGTTTACATACCCCAAGACTTACGCCCTGGTT 48757

Qy 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80

Db 48758 TTGACTCTCGAGAGGATCAAGCTTCCGGATGATGTAATGGCGACATGAAGCTTAGGAGC 48817

Qy 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100

Db 48818 AGTCTAGCCAGAGAGGCGCTTAATCGGTTCTTTTGGCTGGGTTGACCTGGATGGGATGA 48877

Qy 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120

Db 48878 AATCTCACTAGCCCTGTTCAAGCTTCCACGAGTCTGTGGACTCAAAATACGGCGAG 48937

Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140

Db 48938 CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCGCTGCTAAACACCCCTATCGTGA 48997

Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys 155

```
Db 48998 AATTACAGGAAGCAACATCTAGCGCTCTCAAAAAGGAACGG 49042
RESULT 9
AAQ62300/c
ID AAQ62300 standard; DNA; 3382 BP.
XX
XX AAQ62300;
XX
XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 27-MAY-1994 (first entry)
XX
XX pDam-L3 plasmid fragment encoding thermostable ligase.
XX
XX Ligase, thermostable; thermostability; Desulfohalobus ambivalens;
KW archaeobacteria; detection; ligation; ATP; adenosine triphosphate; ss.
XX
XX Acidianus ambivalens.
XX
XX Key Location/Qualifiers
FH 1..273
CDS /tag= a
FT /note= "Open reading frame 4."
FT complement(250..771)
FT /tag= b
FT /note= "Open reading frame 3."
FT 878..883
FT /tag= c
FT 908..2707
FT /tag= d
FT /product= "Ligase."
FT 2713..2722
FT /tag= e
FT /note= "Potential transcription termination signal."
FT 2804..2811
FT /tag= f
FT /note= "Experimental transcription termination signal."
FT 2961..3382
FT /tag= g
FT /note= "Open reading frame 2."
FT
FT
XX DE4217134-A1.
XX
XX 25-NOV-1993.
XX
XX 23-MAY-1992; 92DE-04217134.
XX
XX 23-MAY-1992; 92DE-04217134.
XX
XX (BOE ) BOEHRINGER MANNHEIM GMBH.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Rueger R, Kessler C, Kaletta C, Jarsch M, Kletzin A;
XX
XX WPI: 1993-378402/48.
DR P-PSDB; AAR44614, AAR51076, AAR51077, AAR51078.
XX
XX Thermostable ligase from archaeobacteria - and DNA coding for it, useful
PT for nucleic acid detection.
XX
XX Claim 11; Fig 1; 26pp; German.
XX
XX The DNA ligase is ATP dependent and is useful for detecting nucleic acids
CC by hybridising two oligonucleotides with adjacent sequences of the target
CC nucleic acid, ligating the two oligonucleotides (using the ligase) and
CC detecting the ligation product. (Updated on 25-MAR-2003 to correct PN
CC field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
XX Sequence 3382 BP; 1188 A; 512 C; 683 G; 999 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 3.28e-21 Length: 3382
Score: 242.00 Matches: 58
```

```
Percent Similarity: 57.14% Conservative: 30
Best Local Similarity: 37.66% Mismatches: 52
Query Match: 30.02% Indels: 15
Db: 2 Gaps: 3
US-08-957-709A-71 (1-156) x AAQ62300 (1-3382)
QY 12 IleLeuIleGluProPheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgVal 31
Db 723 ATAGTAATCTCTCTCTTACTCAGATACTATAGAGAGACGGGGTTCGATTTTAGAGTA 664
QY 32 Gly-----ArgGluAlaPheValLysGlyLys----- 40
Db 663 GGAGGAGAGATAGCCCGCTTTTAAGAAGACTGACGAATATATGAAGATGGAAGACCCG 604
QY 41 -----LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAla 58
Db 603 CGTTCAATTTATGAGATAGAGAAAGGAGATGAATTCATCATTTACCCCTAACCAACATGTA 544
QY 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIle 78
Db 543 CTGTTAGTTACTGAGGAGTATGTTAACTGCCGAACGACGTAATGGCTTCGTCACCTTA 484
QY 79 ArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrp 98
Db 483 AGGCTCTCTTTGCCAGGCTGGTCTTTTGTGTACCACCGACTATCGTAGATCGAGGCTTT 424
QY 99 AspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyr 118
Db 423 GAAGGACAACTCACAATAGAAGTACTA---GGTTCTGCTTTTCCCGTGAAGATAAGAGG 367
QY 119 GlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyr 138
Db 366 GGGACTCGCTTTCTTCACCTCATTTTCGCCAGAACATTAAACACCATAGAGATCCTTAT 307
QY 139 ArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 306 CATGCAAAATACCAAGGACA-ACAAGGGCTAACTTTTACCAA 266
RESULT 10
ADK16664
ID ADK16664 standard; DNA; 597 BP.
XX
XX AC ADK16664;
XX
XX 06-MAY-2004 (first entry)
DT
DE Nanoarchaeum equitans cancer-associated (CA) gene #308.
KW cancer-associated gene; CA gene; cancer; carcinoma; lymphoma; leukaemia;
KW ds; gene.
XX
XX Nanoarchaeum equitans.
XX
XX WO2003093434-A2.
XX
XX 13-NOV-2003.
XX
XX 01-MAY-2003; 2003WO-US013699.
XX
XX 01-MAY-2002; 2002US-0377447P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Stetter KO, Waters E, Kretz K, Podar M, Richardson T;
PI Noordewier M;
XX
XX WPI: 2004-053041/05.
DR P-PSDB; ADK16665.
XX
XX New recombinant cancer-associated genes, such as KCNJ9, useful for
PT diagnosing or treating carcinoma, e.g. breast, colon, rectal, pancreatic,
PT cervical, or skin cancers, lymphomas, or leukemia.
PT
```


XX Claim 5; SEQ ID NO 616; 251pp; English.
XX The invention comprises then amino acid and coding sequences of cancer-associated (CA) genes isolated from Nanoarchaeum equitans. The invention CC also comprises the Nanoarchaeum equitans genome. The DNA and protein CC sequences of the invention are useful for diagnosing and treating cancer CC (e.g. carcinoma, lymphoma, or leukaemia). The present DNA sequence CC represents a Nanoarchaeum equitans CA gene of the invention.
XX
SQ Sequence 597 BP; 246 A; 80 C; 104 G; 167 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,45e-20 Length: 597
Score: 227.50 Matches: 63
Percent Similarity: 49.47% Conservative: 31
Best Local Similarity: 33.16% Mismatches: 51
Query Match: 28.23% Indels: 45
DB: 12 Gaps: 8

US-08-957-709A-71 (1-156) x ADK16664 (1-597)

QY 1 MetLeuLeuProAspTrpLysIle-----ArgLysGluIleLeuLeuGlu 15
DB 1 ATGCTATTAAATGATAGAGAAATTAAGAGCTAATAGACAAAAGAGATAATATTAG 60
QY 16 ProPheSerGluGluSerLeuGluProAlaGlyTyrAspLeuArgValGlyArg----- 33
DB 61 CCTTTTTCATGAACAAATCCACCGCATCGATAGATCTAAGGCTAGCAATAGATT 120
QY 34 GluAlaPheValLysGly-----LysLeu 41
DB 121 AGAATTTTCAGAAAGCGCAATATAGAGTAAATTGATCCCAAGGATTTTAAGCATGAAC 180
QY 42 IleAspValGluLys----- 46
DB 181 ATAAGATAGAACAGATGAAATAAATAAATAATTTGAAAAATACAAATATACCGATGTAATT 240
QY 47 -----GluGlyLysValValIleProProArgGluTyrAlaLeuLeuLeuLeu 64
DB 241 ATTACAGAGAACCTTTTCATTATTTATCCAGGGGATTTGTTTAGCCTCTATATATGAA 300
QY 65 ArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeuAlaArg 84
DB 301 TATATAAAATTCGCAAGATATATAGCAGCACAAATTCATGGCAAAATCCTCTATAGCTAGG 360
QY 85 GluGlyValIle-----GlySerPheAlaTTPValAspProGlyTrpAspGlyAsnLeu 102
DB 361 TTAGGTTTAAATATACATACATCTGCAGGTGGATAGACCCCTGTTATAGGGCCATCTA 420
QY 103 ThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGluArgPhe 122
DB 421 ACATTAGAAATATTTAATACTAACAATGCGCAATTAAGCTTTACCTAAATGAAAAATA 480
QY 123 ValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArg----- 139
DB 481 GCACAAATTCGAGCTATTAGAATA---AATCCTGTAGAGAGAGACTATAAAGAAAAAGGA 537
QY 140 GlyAsnTyr-----GlnGlySerThr 146
DB 538 GGTAATAACTATAAAGAAAAAGGTGCTACC 567

RESULT 11

ADK16049_2
Continuation (3 of 5) of ADK16049 from base 200001 (Nanoarchaeum equitans genome.)
WP Sequence split into 5 fragments LOCUS ADK16049 Accession Adk16049

WP	Fragment Name	Begin	End
WP	ADK16049_1	1	110000
WP	ADK16049_2	100001	210000
WP	ADK16049_3	200001	310000
WP	ADK16049_4	300001	400001

Alignment Scores:

Pred. No.: 3,71e-17 Length: 110000
Score: 227.50 Matches: 63
Percent Similarity: 49.47% Conservative: 31
Best Local Similarity: 33.16% Mismatches: 51
Query Match: 28.23% Indels: 45
DB: 12 Gaps: 8

US-08-957-709A-71 (1-156) x ADK16049_2 (1-110000)

QY 1 MetLeuLeuProAspTrpLysIle-----ArgLysGluIleLeuLeuGlu 15
DB 83658 ATGCTATTAAATGATAGAGAAATTAAGAGCTAATAGACAAAAGAGATAATATTAG 83717
QY 16 ProPheSerGluGluSerLeuGluProAlaGlyTyrAspLeuArgValGlyArg----- 33
DB 83718 CCTTTTTCATGAACAAATCCACCGCATCGATAGATCTAAGGCTAGCAATAGATT 83777
QY 34 GluAlaPheValLysGly-----LysLeu 41
DB 83778 AGAATTTTCAGAAAGCGCAATATAGAGTAAATTGATCCCAAGGATTTTAAGGATGAAC 83837
QY 42 IleAspValGluLys----- 46
DB 83838 ATAAGATAGAACAGATGAAATAAATAAATAATTTGAAAAATACAAATATACCGATGTAATT 83897
QY 47 -----GluGlyLysValValIleProProArgGluTyrAlaLeuLeuLeuLeu 64
DB 83898 ATTACAGAGAACCTTTTCATTATTTATCCAGGGGATTTGTTTAGCCTCTATATATGAA 83957
QY 65 ArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeuAlaArg 84
DB 83958 TATATAAAATTCGCAAGATATATAGCAGCACAAATTCATGGCAAAATCCTCTATAGCTAGG 84017
QY 85 GluGlyValIle-----GlySerPheAlaTTPValAspProGlyTrpAspGlyAsnLeu 102
DB 84018 TTAGGTTTAAATATACATACATCTGCAGGTGGATAGACCCCTGTTATAGGGCCATCTA 84077
QY 103 ThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGluArgPhe 122
DB 84078 ACATTAGAAATATTTAATACTAACAATGCGCAATTAAGCTTTACCTAAATGAAAAATA 84137
QY 123 ValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArg----- 139
DB 84138 GCACAAATTCGAGCTATTAGAATA---AATCCTGTAGAGAGAGACTATAAAGAAAAAGGA 84194
QY 140 GlyAsnTyr-----GlnGlySerThr 146
DB 84195 GGTAATAACTATAAAGAAAAAGGTGCTACC 84224

RESULT 12

ADK27081_02/c

Continuation (3 of 17) of ADK27081 from base 200001 (Hyperthermophile Methanopyrus kandi)
WP Sequence split into 17 fragments LOCUS ADK27081 Accession Adm27081

WP	Fragment Name	Begin	End
WP <td>ADK27081_00</td> <td>1</td> <td>110000</td>	ADK27081_00	1	110000
WP <td>ADK27081_01</td> <td>100001</td> <td>210000</td>	ADK27081_01	100001	210000
WP <td>ADK27081_02</td> <td>200001</td> <td>310000</td>	ADK27081_02	200001	310000
WP <td>ADK27081_03</td> <td>300001</td> <td>410000</td>	ADK27081_03	300001	410000
WP <td>ADK27081_04</td> <td>400001</td> <td>510000</td>	ADK27081_04	400001	510000
WP <td>ADK27081_05</td> <td>500001</td> <td>610000</td>	ADK27081_05	500001	610000
WP <td>ADK27081_06</td> <td>600001</td> <td>710000</td>	ADK27081_06	600001	710000
WP <td>ADK27081_07</td> <td>700001</td> <td>810000</td>	ADK27081_07	700001	810000
WP <td>ADK27081_08</td> <td>800001</td> <td>910000</td>	ADK27081_08	800001	910000
WP <td>ADK27081_09</td> <td>900001</td> <td>1010000</td>	ADK27081_09	900001	1010000
WP <td>ADK27081_10</td> <td>1000001</td> <td>1110000</td>	ADK27081_10	1000001	1110000
WP <td>ADK27081_11</td> <td>1100001</td> <td>1210000</td>	ADK27081_11	1100001	1210000
WP <td>ADK27081_12</td> <td>1200001</td> <td>1310000</td>	ADK27081_12	1200001	1310000
WP <td>ADK27081_13</td> <td>1300001</td> <td>1410000</td>	ADK27081_13	1300001	1410000
WP <td>ADK27081_14</td> <td>1400001</td> <td>1510000</td>	ADK27081_14	1400001	1510000
WP <td>ADK27081_15</td> <td>1500001</td> <td>1610000</td>	ADK27081_15	1500001	1610000
WP <td>ADK27081_16</td> <td>1600001</td> <td>1694968</td>	ADK27081_16	1600001	1694968

03-AUG-2000; 2000JP-00280988.

(KYOW) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
Tateishi N, Senoh A, Ikeda M, Ozaki A;
WPI; 2001-376931/40.
P-PSDB; AAG92886.

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analyzing expression profile or pattern of a gene and identifying homologous gene.

Claim 8; SEQ ID NO 3140; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office

Sequence 567 BP; 104 A; 132 C; 165 G; 166 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.43e-18 Length: 567
Score: 214.00 Matches: 56
Percent Similarity: 50.30% Conservative: 29
Best Local Similarity: 33.14% Mismatches: 60
Query Match: 26.55% Indels: 24
DB: 5 Gaps: 4

US-08-957-709A-71 (1-156) x AAH68105 (1-567)

Qy 1 MetLeuLeuProAspTrpLysIleArgLys-----GluIleLeuIleGlu 15
Db 1 GTGCTTCTTCAGATCGTGACATCGTAATCAATTGACGCGGCGACTTGGGAATTGAA 60
Qy 16 ProPheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
Db 61 CCTTTCACGCTGAGCTGATTCAGCCGTCGAGTGTGATGTCGATGCCGATCGACCGCTACTTC 120
Qy 34 GluAlaPheValLysGlyLys----- 40
Db 121 CGGGTTTTCATAATCTCTAAGTAGTACCCACATGACCTAAGTTGAATCAGATGAGCTG 180
Qy 41 -----LeuIleAspValGluLysGluGlyLysValValIleProArgGluTyrAla 58
Db 181 ACAGCCCTTGTGAGGTGTGAGGACGGCGGAGGGCTTGTGTCATCCGGGTGAGTTGTG 240
Qy 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIle 78
Db 241 CTGGCGTCCACGCTGGAAAAGTTCACTTTTCCTGCGCATCTGCCTGCTGTTGGAGGT 300
Qy 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
Db 301 AAGTCGTCTCTGTGTCGTCTTGGGCTGTGTGACGCACTCTACTGCTGGTGTTCATTGATCCT 360
Qy 97 GlyTrpAspGlyAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 116
Db 361 GGTTTGTAGTGTATACATCACAGTGTGAGTGTGTCATATGCGCTAATCTGCCGATCACGTTG 420
Qy 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
Db 421 TGCCCGGGGTATGAGGTGGGCGAGTGTGCTTGTGTCAGATGAGTGTCCCTCGGAGACT 480

CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids, described
CC particularly L-lysine. The present sequence is a nucleic acid, described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the European Patent Office
XX
SQ Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.18e-14 Length: 349980
Score: 214.00 Matches: 56
Percent Similarity: 50.30% Conservative: 29
Best Local Similarity: 33.14% Mismatches: 60
Query Match: 26.55% Indels: 24
DB: 5 Gaps: 4

US-08-957-709A-71 (1-156) x AAH68533 (1-349980)

QY 1 MetLeuLeuProAspTrpLysIleArgLys-----GluIleLeuIleGlu 15
Db 337411 GTGCTTCTTCAGATCGTGACATCGTAATCAATTGACGCGAGCGACTTGGGAATTGAA 337352
QY 16 ProPheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
Db 337351 CCTTTCGACGCTGAGTCAATTCAGCGCTGAGTGTGAGTGTCCGATCGACGCGCTACTTC 337292
QY 34 GluAlaPheValLysGlyLys----- 40
Db 337291 CGGGTTTTCATAACTCTAAGTACACCCACCATTTGACCTTAAGTTGAATCAGGATGAGCTG 337232
QY 41 -----LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAla 58
Db 337231 ACCAGCCTTGTGAGGTTGAGGACGCGAGGGCTTTGTGCTGCATCCGGGTGAGTTTGTG 337172
QY 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIle 78
Db 337171 CTGGCGTCCACCGCTGGAAGTTCACTTTGCCCTGGCATCTGGCTGGTTCGTTGGAGGGT 337112
QY 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
Db 337111 AAGTCGTCTCTGGTTCGCTTGGCTTGTGACGCACTCTACTGCTGGTTTCATTGATCCT 337052
QY 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 337051 GGTTTTAGTGGTTACATCACGTTGGAGTTGTCCAATGTGGCTAATCTGCCGATCACGTTG 336992
QY 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
Db 336991 TGGCCGGGTATGAAGTGGGGCAGTTGGCTTTGTTCCAGATGAGTTCCTCCGCGAGACT 336932
QY 137 ProTyrArgGlyAsnTyrGlnGlySer 145
Db 336931 CCGTATGGTCCGGCAAGCTTGGTTTCG 336905

Search completed: July 2, 2005, 09:23:30
Job time : 763 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 06:11:51 ; Search time 2526 Seconds
(without alignments)
7097.494 Million cell updates/sec

Title: US-08-957-709A-70

Perfect score: 471

Sequence: 1 atgctactccagactggaa.....caagagagaagaactcttag 471

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_est3:*

4: gb_est4:*

5: gb_est5:*

6: gb_est6:*

7: gb_est7:*

8: gb_gse1:*

9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41.6	8.8	712	4	BJ624188
2	40.4	8.6	1101	9	CNS0039G
3	39.6	8.4	756	9	AG530407 Mus muscu
4	39	8.3	336	2	BF281615 EST446206
5	38	8.1	714	9	AG318648 Mus muscu
6	37.8	8.0	1101	9	CNS0106X
7	37.6	8.0	604	9	CW017266 ZMMBLC001
8	37.6	8.0	808	9	AG404615 Mus muscu
9	37.6	8.0	1101	9	CNS0006J
10	37	7.9	1019	9	CNS0107N
11	36.8	7.8	997	9	CNS005TE
12	36.6	7.8	409	9	CE439533
13	36.4	7.7	577	8	AQ416995
14	36.2	7.7	841	7	CK196609 FGAS00506
15	36	7.6	519	8	BZ915075 CH240_57D
16	36	7.6	522	4	BI501756 tm04f06.y
17	36	7.6	604	4	BJ106515
18	36	7.6	613	4	BJ770456
19	36	7.6	661	7	CK580840 IST.W15.2
20	36	7.6	893	5	BQ139899 NF026C1IP
21	36	7.6	973	4	BG261309 602373061
22	35.8	7.6	703	9	AG323494 Mus muscu
23	35.8	7.6	1101	9	CNS00L72
24	35.6	7.6	857	9	CR247912 Forward s

C 25	35.4	7.5	676	7	CN824978
C 26	35.4	7.5	681	9	CL633465
C 27	35.4	7.5	746	8	BH541374 BOHGX75TF
C 28	35.4	7.5	1204	9	CNS016E2
C 29	35.2	7.5	725	6	CA138200 SCEPRT204
C 30	35.2	7.5	1009	5	BX405692 BX405692
C 31	35.2	7.5	2275	1	AF034173
C 32	35	7.4	665	9	CE285308 tigr-g98-
C 33	34.8	7.4	803	9	CW017844 ZMMBLC001
C 34	34.8	7.4	951	8	CC367157 PUHTL06TB
C 35	34.6	7.3	467	1	AI825671 wb75e03.x
C 36	34.6	7.3	664	5	BQ407732 GA_EG000
C 37	34.6	7.3	677	9	CL933212 OA_ABA004
C 38	34.6	7.3	721	2	BF278028 GA_EB003
C 39	34.6	7.3	961	8	CC461360 ZMMBLC035
C 40	34.6	7.3	1100	9	CNS014RI
C 41	34.6	7.3	1164	9	CL048873
C 42	34.4	7.3	516	8	AQ571364 HS_5376.A
C 43	34.2	7.3	277	6	CD342372 EtEstee51
C 44	34.2	7.3	492	9	CE043118 tigr-g98-
C 45	34.2	7.3	538	6	CB472434 sn55_B07.

ALIGNMENTS

RESULT 1
BJ624188 712 bp mRNA linear EST 01-OCT-2003
LOCUS BJ624188 NIBB Mochii normalized Xenopus early gastrula library
DEFINITION Xenopus laevis cDNA clone XL207j03 5', mRNA sequence.
ACCESSION BJ624188
VERSION BJ624188.1 GI:37267690
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.

REFERENCE 1 (bases 1 to 712)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.
http://xenopus.nibb.ac.jp.

FEATURES
source Location/Qualifiers
1..712
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL207j03"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN
Query Match 8.8%; Score 41.6; DB 4; Length 712;
Best Local Similarity 51.6%; Pred. No. 0.39;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
Qy 15 CTGGAATCAGAAAGAAATCTTATAGCCATTTCTGAGATCGCTCCACGCG 74
Db 178 CTGGAGAAACAGACGACGAGACTTTATGGAGCAAGAGACGCGATCTCCAGACATT 237

```
QY 75 AGGTTATGACCTCAGAGTGGGAGAGAGCGCTTTTGTAGGGGAAATTAATCGAGTGA 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 AACAAATGGGACACAGCTGGCTACAGAGGATTTATATCCAGAGAATTTGATAAGAGGGA 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 135 AAAGGAAGGAAAGTGGTATTCTCCTCAAGGGAATACGCCCTTAATCCTAACCTCGAGAG 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 AGCATCGATCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 195 GATA 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 GATA 361

RESULT 2
CNS0039G/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BAC08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL063921.1 GI:4941778
VERSION
AL063921.1
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoler in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and Est libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BAC08K10"
/clone_lib="RPCI-98"
/notes="end : TET3"
source

ORIGIN
Query Match 8.6%; Score 40.4; DB 9; Length 1101;
Best Local Similarity 13.5%; Pred. No. 1;
Matches 52; Conservative 182; Mismatches 150; Indels 0; Gaps 0;

QY 21 AATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAAATCGCTCAACCCAGCAGTTA 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1077 WTTTWWKDRADRRWAGDADRWWDGAGTWTATATTTTWWTATTTTWWTATTTTWWTATAAK 1018

QY 81 TGACCTCAGAGTGGGAGAGAGCGCTTTTGTAGGGGAAATTAATCGAGTGAAGGA 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 TDTAATWTRADWAGDRGAGKRDRDAATDADGAGRRDGGKRKRDGKDDDDDK 958

QY 141 AGGAAAGTGGTATTCTCCTCAAGGGAATACGCCCTTAATCCTAACCTCGAGAGGATAA 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 957 GGGKKKAAAKATKATKWDMDWDKDWKDGAKDRKADDDGAGDKDDGKGDADDDTD 898
QY 201 GTTGGCCGACGATGTTATGGGGATATGAAGATAAGGAGACGCTTTAGCAAGAGAAGGGGT 260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 897 GTKDDDDKDKWDDWDKAGTGWGADATWAAATDWWWWGADADWWTWDAADDDWADRDWD 838
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 261 TATTGGTCTTTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 AAWKWDADAWAGTADRDWDGDRAGKRGGRKRRDRKRDADDDRDADAATTTWT 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 321 CAATCCCTCAATGAACCTGTGCAATTAAAGATATGAGAGAGATTTGTCAGATCGCAT 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 TTTTTRDDDKWKTDTWTRWAADRTDRTDRTDRTDRTDRTDRTDRTDRTDRTDRTDRT 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 381 TATAAGGCTAGAGGTCGGCAAG 404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 717 DDADADDATDRTDRTDRTDRTDRTDRTDRTDRTDRTDRTDRTDRTDRTDRTDRTDRT 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
AGS30407/c
LOCUS
DEFINITION
Mus musculus molossinus DNA, clone:MSMg01-440D22.T7, genomic survey
sequence.
ACCESSION
AGS30407
AGS30407.1 GI:48290795
VERSION
AGS30407
KEYWORDS
GSS.
SOURCE
Mus musculus molossinus
ORGANISM
Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
BAC end Sequences of Library MSMg01
2 (bases 1 to 756)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center, The Institute of Physical and
Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
Location/Qualifiers
1..756
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-440D22.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
source

ORIGIN
Query Match 8.4%; Score 39.6; DB 9; Length 756;
Best Local Similarity 52.4%; Pred. No. 1.6;
Matches 87; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 218 TGGGGGATAGAAGATAGGAGCAGTTTACGAGAGAGGGGTTATTGTTCTTTGCTT 277
```

Db 395 TGGAGTGGATGGTATTGAGCAGGGAAGGGGGAAGGATAGGGGGGGTTCCTAA 336
Qy 278 GGGTTGACCCAGGATGGATGGAACCTTAACACTTAATGCTCTACAATGCCTCAAAATGAAC 337
Db 335 GGGTAAACACGGAAGGGGATACATTTAAATGTAATAACGAAATATCTAATAAAAA 276
Qy 338 CTGTCGAATTAAGATATGGAGAGAGATTTGTGCGATGCGATTAT 383
Db 275 ATGTAGATTTTCAGATATGGGAAATTAGATAGTGAATTTCTTTGT 230

RESULT 4
BF281615/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BF281615 336 bp mRNA linear EST 28-NOV-2000
EST446206 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
Rattus norvegicus cDNA clone RGIAN54, mRNA sequence.
BF281615
EST.
BF281615.1 GI:11212685
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 336)
Malek, R.L., Cho, J., Lee, Y., Karamycheva, S., Parvizi, B., Pertea, G.,
Sultana, R., Tsai, J., White, J., Quackenbush, J. and Lee, N.H.
Generation of ESTs from Normalized Rat Embryo, Bento Soares
Unpublished (2000)
Other_ESTs: EST446205 EST446207
Contact: Lee, NH

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.

FEATURES
source
1..336
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="RGIAN54"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/notes="Vector: pT377pac; Site 1: EcoRI; Site 2: NotI;
Combination of ROV, RBR, RKI, RLI, RPL, RLU, REM, RMI,
RSP, RHE, RPC, RPN"

ORIGIN
Query Match 8.3%; Score 39; DB 2; Length 336;
Best Local Similarity 52.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
Qy 90 AGTGGCAGAGAGCTTTCTTAAAGGGAATTAATCGACGTGGAAAGCAAGGAAAGT 149
Db 335 AGTAGGCGAGCTGTCTGCTCTCAAAAAGGGAAGGGAAGGAGGAGGAGGGGT 276
Qy 150 CGTTATTCCTCCAAAGGGAATACGCTTAATCCCTAACCCCTCGAGAGGATAAAGTTGCCCA 209
Db 275 CATGAGTCATGACCAACATCATCAATTAATTAATCATGTAACAGATATTTGAAAAA 216
Qy 210 CGATGTTATGGGGATATGAAGTAAGGAGCAGTTTAGCAAGAGAG 256
Db 215 AAAAAATTAAAGCATAAGGAAATGAAAAGGATTTAGGTACAAAAG 169

RESULT 5
AG318648/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AG318648 714 bp DNA linear GSS 02-JUN-2004
Mus musculus molossinus DNA, clone:MSMg01-103007.T7, genomic survey
sequence.
AG318648
AG318648.1 GI:47891605
GSS.
Mus musculus molossinus
Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
BAC end Sequences of Library MSMg01
Unpublished
2 (bases 1 to 714)
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.

FEATURES
source
1..714
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-103007.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 8.1%; Score 38; DB 9; Length 714;
Best Local Similarity 51.8%; Pred. No. 4;
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
Qy 218 TGGGGGATATGAAGATAAGGAGCAGTTTACGAAGAGAGGGTTATTGGTTCTTTTCTT 277
Db 388 TGGGAGTGGATGGGTATTGAGCAGGGGAGGGGAGGATAGGGGGGGTTCCTAA 329
Qy 278 GGGTTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCTCAAAATGAAC 337
Db 328 GGGTAAACACGGAAGGGGATACATTTAAATGTAATAACGAAATATCTAATAAAAA 269
Qy 338 CTGTCGAATTAAGATATGGAGAGAGATTTGTGCGATGCGATTAT 383
Db 268 ATGTAGATTTTCAGATATGGGAAATTAGATAGTGAATTTCTTTGT 223

RESULT 6
CNS0106X
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS

CNS0106X 1101 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL098595
AL098595.1 GI:5610206
GSS.

SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.
FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN03K20"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : T7"
ORIGIN
Query Match 8.0%; Score 37.8; DB 9; Length 1101;
Best Local Similarity 15.2%; Pred. No. 6.1; Mismatches 128; Indels 0; Gaps 0;
Matches 50; Conservative 151
QY 135 AAAGGAAGGAAAGTGGTTATTCCTCCAAAGGAATACGCTTAATCTTAACCTCGAGAG 194
DB 749 AKADARDKRRWDKDKRKAARAKAADDADADAKADAKADAKADADADDGRGDG 808
QY 195 GATAAGTGGCCGACGATGTTATGGGGATATGAAGATAAGGAGCAGTGTAGCAGAGA 254
DB 809 KKKRAKDRKKKKKKKANGDKKKAADAKADAKADAKADAKADAKADAKADAKA 868
QY 255 AGGGGTATGTTGCTTTGCTGGGTGACCCAGGATGGAACTTAACTTAACACTAAT 314
DB 869 AKKAKADDDDAKAAATKAAKATKAKDKKAKAKKKKKKKDKAKAKADKDKDD 928
QY 315 GCTCTACAATGCTCAAACTGAACCTGTGCAATTAAGATATGAGAGAGATTTGTGCAGAT 374
DB 929 KDKKDAKADKKKKKDKRAKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDD 988
QY 375 CGCATTTATAAGCTTAGAGGTCCGGCAAGAAACCTTTACAGAGGAACCTATCAGGGGAG 434
DB 989 AKAKAKAADDAAADAKAADAADAKAKADADAAKAKADADAAKADADADAKAKAAKRAAK 1048
QY 435 CACNAGTTAGCGTTTCAAGAGAAGA 463
DB 1049 AKADADAADAKADKDKDKADKDDADKA 1077
RESULT 7
CW017266/c 604 bp DNA linear GSS 23-SEP-2004
LOCUS ZMMLC0010B24.r ZMMLC Zea mays genomic clone ZMMLC0010B24 3',
DEFINITION genomic survey sequence.
ACCESSION CW017266
VERSION CW017266.1
KEYWORDS GI:52607659
SOURCE GSS.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 604)

AUTHORS Wing, R., Luo, M., Soderlund, C. and Haller, K.
TITLE ZMMLC sequences
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
Plate: 0010 row: B column: 24
Class: BAC ends.
FEATURES
Location/Qualifiers
1..604
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMLC0010B24"
/tissue type="immature ears"
/lab host="DH10B T1 phage resistant"
/clone_lib="ZMMLC"
/note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN
Query Match 8.0%; Score 37.6; DB 9; Length 604;
Best Local Similarity 47.1%; Pred. No. 6;
Matches 115; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 118 AAATTAATCGACGTGGAAAGGAAGAAAGTCGTTATTCCTCCAAAGGAATACGCTTA 177
DB 479 ATAATACCAAGTTGTAAACTATAGAGAAGGGGATACTCTCTCATGTACTTTGTCTCC 420
QY 178 ATCCTAACCTCGAGAGGATAAGTTGCCGACGATGTTATGGGGATATGAAGATAAGG 237
DB 419 ATTTATTATCATGTGAATACCAACCCAAAAGTGGCTAATTAACGAATGAGACATGA 360
QY 238 AGCAGTTTAGCAAGAAGAGGGGTATTTGGTTCTTTTGTGTTGGTTCACCCAGGATGGGAT 297
DB 359 ACNACTTTGGCATCATGCTGGAGTTTATATGATTTGTCATTCCTCCTAAAGATAGTAT 307
QY 298 GGAAGACTTAACACTAATGCTCTACAATGCCTCAATGAACCTGTGCAATTAAGATATGA 357
DB 299 AAAAGAAATATACAAATGTTGAATAGGCATTAACTTAATTTGGAAAATTAAGAATTGA 240
QY 358 GAGA 361
DB 239 GAAA 236
RESULT 8
AG04615/c 808 bp DNA linear GSS 03-JUN-2004
LOCUS Mus musculus molossinus DNA, clone:MSMg01-262116.T7, genomic survey
DEFINITION sequence.
ACCESSION AG04615
VERSION AG04615.1
KEYWORDS GI:48047301
SOURCE GSS.
ORGANISM Mus musculus molossinus
Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
BAC end Sequences of Library MSMg01
2 (bases 1 to 808)
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/.

COMMENT
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSNg01. For BAC library availability, please contact Kuniya Abe (abe@c.riken.jp). Tsukuba Institute, Bio Resource Center, The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@c.riken.jp

PRIMERS

Sequencing : T7

LIBRARY

Vector : pBACe3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI.

Location/Qualifiers

1. .808

/organism="Mus musculus molossinus"

/mol_type="genomic DNA"

/sub_species="molossinus"

/db_xref="taxon:57486"

/clone="MSNg01-262116.T7"

/sex="male"

/tissue type="mixture of kidney and spleen"

/clone_lib="MSNg01 Mouse Male BAC Library"

ORIGIN

Query Match 8.0%; Score 37.6; DB 9; Length 1101;
Best Local Similarity 54.3%; Pred. No. 6.4;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 12 AGACTGGAATCAGAAAGAAATCTATTAGAGCCATTTCTGAAGATCGCTCAACC 71
|||||
DB 699 AGATAGGAGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 640
QY 72 AGCAGGTATGACCTCAGAGTGGCAGAGAGGCTTTTGTAAAGGGAATTAATCGAGT 131
|||||
DB 639 AAAAGAAATATAAGAGGGGGAAGAGGAGTTGATTAATTAAGTATAAGAT 580
QY 132 GGAAGAGGAGGAAAGTCG 151
|||||
DB 579 GAAAGAAAGTAAATTAG 560

Query Match 8.0%; Score 37.6; DB 9; Length 808;

Best Local Similarity 54.3%; Pred. No. 6.4;

Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 12 AGACTGGAATCAGAAAGAAATCTATTAGAGCCATTTCTGAAGATCGCTCAACC 71

DB 699 AGATAGGAGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 640

QY 72 AGCAGGTATGACCTCAGAGTGGCAGAGAGGCTTTTGTAAAGGGAATTAATCGAGT 131

DB 639 AAAAGAAATATAAGAGGGGGAAGAGGAGTTGATTAATTAAGTATAAGAT 580

QY 132 GGAAGAGGAGGAAAGTCG 151

DB 579 GAAAGAAAGTAAATTAG 560

RESULT 9

CNS0006J

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR01M22 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL062049

AL062049.1

GI:4938511

GSS.

Drosophila melanogaster

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999)

Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammeter in Peter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

Location/Qualifiers
1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR01M22"
/clone_lib="RPI-98"
/note="end : TET3"

ORIGIN

Query Match 8.0%; Score 37.6; DB 9; Length 1101;
Best Local Similarity 29.4%; Pred. No. 7;
Matches 131; Conservative 94; Mismatches 218; Indels 3; Gaps 1;

QY 19 AAATCAGAAAGAAATCTATTAGAGCCATTTCTGAAGATCGCTCAACAGCAGGT 78
|||||
DB 489 AAAAARARAAATGARAATRWARAGRTTATTTGAAAAAATAAAATTTTAAAGAAAAA 548
QY 79 TATGACCTCAGAGTGGCAGAGAGGCTTTTGTAAAGGGAATTAATCGACGTGGAAG 138
|||||
DB 549 AAATATATAAATTTTTRRRTRRAGTATTTTTRAKAAAAAATAATTTAGAAAAAG 608
QY 139 GAAGGAAAGTCTTATTCTTCAAGGGAATAGCGCTTAATCTCAACCTCGAGAGGATA 198
|||||
DB 609 AAAAARARAAAGTCTTTGAAGAAAAAATAAATAAATAAATAAATAAATAAATAA 668
QY 199 AGATTGCCGAGATTTATGGGGATATGAAGATGAAGAGCAGTTTAGCAAGAGAGGG 258
|||||
DB 669 AATAAAAAAAAWRTCKTATAAAAAATAAAGAGGAGGAGGAGGAGGAGGAGG 728
QY 259 GTTATTGGTCTTTTCTTGGTTGACCCAGGATGGATGGAACTTAACACTTAATGCTC 318
|||||
DB 729 AWRRAAAAKTKRKKTKTKRKAAGRR---ARRAWAGAAAAAATAAATAAATAATTKGTWAKA 785
QY 319 TACAATGCTCAATGAACCTGTGGAATTAAGATATGGAGAGAGATTTGTGCAGATCGCA 378
|||||
DB 786 AWRRAAAAKTKRKKTKTKRKAAGRR---ARRAWAGAAAAAATAAATAAATAATTKGTWAKA 845
QY 379 TTTATAAGCTAGAGGTCGGCAAGAAACCTTACAGAGGAACTATCAGGGAGCACA 438
|||||
DB 846 KGKDGTRARRRAGAGDGDWDAKAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 905
QY 439 AGTTAGCTTTTCAAGAGAGAAAGAA 464
|||||
DB 906 RDKDKRRAGWGTRRRRARARTRKA 931

RESULT 10

CNS0107N/c

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL098621

AL098621.1

GI:5610232

GSS.

Drosophila melanogaster

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1019)

Genoscope.

Direct Submission

Submitted (23-JUL-1999)

Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammeter in Peter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

COMMENT

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

Drosophila melanogaster genome survey sequence Sp6 end of BAC BACR03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 07:12:26 ; Search time 484 Seconds
(without alignments)
6103.716 Million cell updates/sec

Title: US-08-957-709A-70
Perfect score: 471
Sequence: 1 agctacttccagactgga.....caagagaagaactctag 471

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/PCTUS_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.2	8.5	2964	20	US-10-784-986-3
2	38.6	8.2	89047	18	US-10-672-787-34
3	37.2	7.9	110021	21	US-10-461-862-83
4	36.2	7.7	644	13	US-10-027-632-224178
5	36.2	7.7	644	17	US-10-027-632-224178
6	35	7.4	367378	16	US-10-312-841-1
7	34.8	7.4	570	9	US-09-864-761-7422

8	34.8	7.4	570	9	US-09-864-761-8565	Sequence 8565, Ap
9	34.6	7.3	649	13	US-10-027-632-226698	Sequence 226698,
10	34.6	7.3	649	17	US-10-027-632-226698	Sequence 226698,
11	34.6	7.3	90435	19	US-10-322-696-160	Sequence 160, App
12	34.2	7.3	449	14	US-10-123-155-224	Sequence 224, App
13	34.2	7.3	449	15	US-10-146-731-224	Sequence 224, App
14	34.2	7.3	449	15	US-10-140-472-224	Sequence 224, App
15	34.2	7.3	449	15	US-10-141-761-224	Sequence 224, App
16	34.2	7.3	449	16	US-10-142-885-224	Sequence 224, App
17	34.2	7.3	449	16	US-10-158-790-224	Sequence 224, App
18	34.2	7.3	449	17	US-10-137-871-224	Sequence 224, App
19	34.2	7.3	449	17	US-10-140-923-224	Sequence 224, App
20	34.2	7.3	449	17	US-10-141-756-224	Sequence 224, App
21	34.2	7.3	449	17	US-10-141-759-224	Sequence 224, App
22	34.2	7.3	449	17	US-10-140-805-224	Sequence 224, App
23	34.2	7.3	449	17	US-10-140-864-224	Sequence 224, App
24	34.2	7.3	449	18	US-10-142-426-224	Sequence 224, App
25	34	7.2	1691139	14	US-10-067-514-1	Sequence 1, Appli
26	34	7.2	1691139	17	US-10-419-723-1	Sequence 1, Appli
27	33.8	7.2	2256646	19	US-10-470-565-1	Sequence 1, Appli
28	33.6	7.1	405	10	US-09-975-719-242	Sequence 240, App
29	33.6	7.1	765	10	US-09-975-719-240	Sequence 240, App
30	33.6	7.1	12269	15	US-10-311-455-1384	Sequence 1384, Ap
31	33.6	7.1	12269	18	US-10-221-714A-194	Sequence 194, App
32	33.6	7.1	38844	13	US-10-060-333-3	Sequence 3, Appli
33	33.6	7.1	42235	10	US-09-975-719-1	Sequence 1, Appli
34	33.4	7.1	700	17	US-10-398-221-2976	Sequence 2976, Ap
35	33.4	7.1	746	20	US-10-425-115-20392	Sequence 20392, A
36	33.4	7.1	2107	20	US-10-739-930-4920	Sequence 4920, Ap
37	33.4	7.1	114931	13	US-10-087-192-862	Sequence 862, App
38	33.2	7.0	985	18	US-10-425-114-6662	Sequence 6662, Ap
39	33.2	7.0	1945	20	US-10-425-115-101174	Sequence 101174,
40	33	7.0	326	11	US-09-732-627A-4547	Sequence 4547, Ap
41	33	7.0	2731748	19	US-10-297-465A-1	Sequence 1, Appli
42	32.8	7.0	544	14	US-10-184-644-274	Sequence 274, App
43	32.8	7.0	544	14	US-10-184-634-274	Sequence 274, App
44	32.8	7.0	775	14	US-10-123-155-120	Sequence 120, App
45	32.8	7.0	775	15	US-10-146-731-120	Sequence 120, App

ALIGNMENTS

RESULT 1
US-10-784-986-3
; Sequence 3, Application US/10784986
; Publication No. US20040229311A1
; GENERAL INFORMATION:
; APPLICANT: HIRANO, Seiko
; APPLICANT: YASUEDA, Hisaehi
; TITLE OF INVENTION: Novel lysine decarboxylase gene and method for
; TITLE OF INVENTION: producing L-lysine
; FILE REFERENCE: US-109
; CURRENT APPLICATION NUMBER: US/10/784,986
; PRIOR FILING DATE: 2004-02-25
; PRIOR APPLICATION NUMBER: JP 2003-47185
; PRIOR FILING DATE: 2003-02-25
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2964
; TYPE: DNA
; ORGANISM: Methylophilus methylotrophus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (684)..(2930)
US-10-784-986-3

Query Match 8.5%; Score 40.2; DB 20; Length 2964;
Best Local Similarity 46.3%; Pred. No. 0.091;
Matches 132; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 87 CAGAGTGGCAGAGAGCGCTTTTGTAAAGGGGAAATTAATCGACGTGGAAAGGAAA 146

```

; Sequence 83, Application US/10461862
; Publication No. US20050090434A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001800
; CURRENT APPLICATION NUMBER: US/10/461,862
; CURRENT FILING DATE: 2003-06-13
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 110021
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(110021)
; OTHER INFORMATION: n = A,T,C or G
US-10-461-862-83

Query Match      7.9%; Score 37.2; DB 21; Length 110021;
Best Local Similarity 52.6%; Pred. No. 4.7;
Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 143 GAAAGTCTGTTATTCCTCCAAAGGAATACGCGCTTAATCCTAACCCCTCGAGAGGATAAAGT 202
DB 81964 GGACATTCATCAGTACCACTTAGAATCTTTCACCCCTGACCATGAGATGAATA 81905

QY 203 TGCCCGCAGATGTTATGGGGATATGAAGATAAGGAGCAGTTTACCAAGAGAGGGGTTA 262
DB 81904 GGCCACACCGTTATGTTTTCTTAAATAAGGAGCGATTGAGAAAACAGAACAGATCAT 81845

QY 263 TTGGTCTTTTGGTGGTTGCGTTGACCCAGCATGGGA 296
DB 81844 TTCTGTCCTTTTCAGTGAGGAAGCTGAGGTAGGTA 81811

RESULT 4
US-10-027-632-224178
; Sequence 224178, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 224178
; LENGTH: 644
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-224178

Query Match      7.7%; Score 36.2; DB 13; Length 644;
Best Local Similarity 53.1%; Pred. No. 0.86;
```

```

195 CATAGTTGACCCCAAGCAATTGACCCGAGTCGTTTTCGAGGTCTCCGCAAGGCTA 254
147 AGTCGTTATCTCTCAAGGGAATACGCTTAACTTAACCTTCGAGAGGATAAAGTTGCC 206
255 TTGCGTGAATTCCTCCCTAACTATTTGCACCTGGCGGCACGGTAGAGATATTTCCGTAATCC 314
207 CGACATGTTATGGGGGATATCAAGATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGG 266
315 TCGCTCTGTACTGACTGTATGCTCGGCAAGTCGACTTATGCGCGTTGCGGCATTATCGT 374
267 TTCTTTTGTGTTGGTTTGCACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGC 326
375 CAACGTCACCCCTTTGACCAAGAGTGGGAGGCTATGTCACTAGAGTTTCAGCAACAC 434
327 CTCAAATGAACCTGTGCAATTAAGATATGGAGAGAGATTGTGCA 371
435 CACACCGCTACCGCCAAATTTATGCTGGCGAAGGCTGTGGCA 479

RESULT 2
US-10-672-787-34
; Sequence 34, Application US/10672787
; Publication No. US20040067554A1
; GENERAL INFORMATION:
; APPLICANT: LAGACE, Robert, E.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BERG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: ELIURA.025C1
; CURRENT APPLICATION NUMBER: US/10/672,787
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 09/596,002
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 34
; LENGTH: 89047
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 66874
; OTHER INFORMATION: a or g or c or t, unknown, or other
US-10-672-787-34

Query Match      8.2%; Score 38.6; DB 18; Length 89047;
Best Local Similarity 47.3%; Pred. No. 1.5;
Matches 116; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 152 TTATCTCTCAAGGGAATACGCTTAACTTAACCTTCGAGAGGATAAAGTTGCCGAGC 211
DB 57078 TCATTCACCAAAATTCATTTGCTTTGGCGGCACCTGTTGAGTATTTTCGCATCCCAAGAG 57137

QY 212 ATGTTATGGGGATATGAAGATAAGGAGCAGTTTAGCAGAGAGGGGTTATTGTTCTT 271
DB 57138 ATGTACTGACGATCTGTTTGGGCAATCCACTTATGACGCTGTGGAATCATGTCAATG 57197

QY 272 TTGCTTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGCTCAA 331
DB 57198 TCACGCCACTTGAGCTGTAATGGGAGGGGCGATGTAACCTTAGAATTTAGCAATACCACA 57257

QY 332 ATGAACCTGTGCAATTAAGATATGGAGAGAGATTTGTGAGATCGCATTTTATAAGGCTAG 391
DB 57258 ATTTGCGACGACGCTATCTATGCTGTGAGGGGTGTGGCAGACAGATGCTATTTTCCCAATCAG 57317

QY 392 AGGCT 396
DB 57318 ATGCT 57322

RESULT 3
US-10-461-862-83/c
```

Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 4 CTAATCCAGCTGGAAATCAGAAAGAATACTTATAGAGCCATTTCTGAAGATCG 63
Db 432 CTCAAACTAAATCGAGATATTAAAGAGGAAAGGATTAATCCCATCTTTGGTAGAAGG 491
QY 64 CTCACACCGAGCGTTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGGAATTA 123
Db 492 CTTCAAGCAGTACAGATAAGCCCTTAAAGGTGCTTAGCATCCTTTTAGATAAGGAATATA 551
QY 124 ATCGACGTGGAAAGGAAAGGAAAG 148
Db 552 ATAGATGTCACAGATAAGGAAAG 576

RESULT 5
US-10-027-632-224178
; Sequence 224178, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 224178
; LENGTH: 644
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-224178

Query Match 7.7%; Score 36.2; DB 17; Length 644;
Best Local Similarity 53.1%; Pred. No. 0.86;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 4 CTAATCCAGCTGGAAATCAGAAAGAATACTTATAGAGCCATTTCTGAAGATCG 63
Db 432 CTCAAACTAAATCGAGATATTAAAGAGGAAAGGATTAATCCCATCTTTGGTAGAAGG 491
QY 64 CTCACACCGAGCGTTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGGAATTA 123
Db 492 CTTCAAGCAGTACAGATAAGCCCTTAAAGGTGCTTAGCATCCTTTTAGATAAGGAATATA 551
QY 124 ATCGACGTGGAAAGGAAAGGAAAG 148
Db 552 ATAGATGTCACAGATAAGGAAAG 576

RESULT 6
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO

; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 7.4%; Score 35; DB 16; Length 3673778;
Best Local Similarity 59.6%; Pred. No. 1.2e+02;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 208 GACGATGTTATGGGGATATGAGATAAGGACAGTTTACGACAGAGAGGGTTATTGGT 267
Db 3478762 GAAGTCGTTTGGGAAATTTGAAGATATTGGGATATTTAGAAGGTTAAGGGGATTTAGTT 3478821
QY 268 TCTTTTGTGTTGGTTCACCCAGGATGGGATGGAAACTTA 306
Db 3478822 TATTTTGGGTAAAGTGTGGGAGTGAAGGTAGGTAA 3478860

RESULT 7
US-09-864-761-7422
; Sequence 7422, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21

```

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7422
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018367.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-7422

```

Query Match	7.4%	Score 34.8	DB 9	Length 570
Best Local Similarity	47.0%	Prod. No. 2.3		
Matches 108	Conservative 0	Mismatches 122	Indels 0	Gaps 0
Qy	155	TTCTCTCAAGGGAATACGCCTTAATCTCTAAACCTCGAGAGGATAAAGTTGCCGCCGACGATG	214	
Db	229	TGCCACGAGGAGACTCAACACAAATCCCATCAAACTGGATGTTAAGATTGCCACCTGGAC	288	
Qy	215	TTATGGGGGATATGAAGATATAGGAGCAGTTTAGCAAGAGAGAGGGGTTATTGGTTCCTTTTG	274	
Db	289	ACTTCGTGCTCTTCTCATATTAAAGTCAACAGGCTTAAAGAGGGAGTTACAGTGTGTGGCTA	348	
Qy	275	CTTCGGTGTGACCCAGGATGGGATGGAACTTTAAACACTAATGCTCTFACAATGCTCTCAAAATG	334	
Db	349	AGGTGATTGACCCAGACTATCATGATGAATCAGCCTACTACTCCACAATGAGGTAAAG	408	
Qy	335	AACCTGTGGAATTAAAGATATGGAGAGAGATTGTGTCAGATCGCATTTATA	384	
Db	409	AAGAGTACGCATGGAAATACAGAGATCCATTAGGGCATCTCTTAGTATTA	458	

RESULT 8
US-09-864-761-8565
; Sequence 8565, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeo mica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8565
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL157957.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
;
US-09-864-761-8565

Query Match 7.4%; Score 34.8; DB 9; Length 570;
Best Local Similarity 47.0%; Pred. No. 2.3;
Matches 108; Conservative 0; Mismatches 122; Indels 0; Gaps 0

Qy 155 TTCCTCAAGGGAATAGCGCTTAATCTCAACCCTCGAGAGAGATAAAGTTGCCCGACGATG 214
Db 229 TGCACGAGGAGACTCAACAACAATCCATCAAACTGCGATGTTAAGATTGCCACCTGGAC 288
Qy 215 TTATGGGGGATATGAGATAAGNGCAGTTTAGCAAGAGAGGGGTATTGTTCTTTTG 274
Db 289 ACTTCGTGCTCTTCCTACATTTAAGTCAACAGGGCTAAAGAGGAGTTACAGTGTGGCTA 348
Qy 275 CTTGGGTTGACCCAGGATGGATCGAAACTTAAACACTAATGCTCTACAAATGCTCAAATG 334
Db 349 AGGTGATTGACCCAGCATATCATGATGAATCAGCCTACTACTCCACAAATGAAGGTAAAG 408
Qy 335 AACCTGTGCAATTAAGATATATGGAGAGAGATTTGTGCAGATCGCAATTTATA 384
Db 409 AAGAGTATCGCATCGAATACAGGACATCCATTAGGGCATCTCTTAGTATTA 458

```

```

RESULT 9
US-10-027-632-226698
; Sequence 226698, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363

```


Db 202 PRATLATLATRAQTAVATTANTSSPMSTRPSKHPMSDTAASFPVPPMRPQAQGPISQVS 261
QY 303 CTTAACACT 311
Db 262 VDQPVVNTT 270

RESULT 15
US-10-141-761-224
; Sequence 224, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P33301C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 224
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-761-224

Query Match 7.3%; Score 34.2; DB 15; Length 449;
Best Local Similarity 20.5%; Pred. No. 3.2;
Matches 51; Conservative 65; Mismatches 133; Indels 0; Gaps 0;
QY 63 GCTCCACACGACGAGTTATCACCACGAGTGGCGACAGAGGCTTTTGTAAAGGGAAATT 122
Db 22 SNDPRNFVPMKMWGLVKRNASVETVDNKTSEDVTMAAASPVLTTRKTSAAHLNSMEVTT 81
QY 123 AATCGAGCGTGAAGAAAGTGGTTATTCTCCAAAGGAATACGCCTTAATCCT 182
Db 82 EDTSRDVSPEATSGVAADGVTIAPTAVASSTTAASMTVASSAPTAASTT 141
QY 183 AACCCCTGAGAGGATAAGTGGCCGACGATGTTATGGGGATATGAAGATAAGAGCAG 242
Db 142 VASIAPTTAASMTAASSTPMTLALPAPTSTSTGRTPSTTATGHPSLTALAQVPKSSAL 201
QY 243 TTATGCAAGAGAGGGTTATGCTTTTGTGCTGGTGGTACCAGGATGGATGAAA 302
Db 202 PRATLATLATRAQTAVATTANTSSPMSTRPSKHPMSDTAASFPVPPMRPQAQGPISQVS 261
QY 303 CTTAACACT 311
Db 262 VDQPVVNTT 270

Search completed: July 2, 2005, 09:10:49
Job time : 493 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 06:11:49 ; Search time 145 Seconds
(without alignments)
5315.078 Million cell updates/sec

Title: US-08-957-709A-70

Perfect score: 471

Sequence: 1 atgctactccagactggaa.....caagagagaagaactcttag 471

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 240568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	471	100.0	471	4	US-09-399-003-70
2	316	67.1	740	3	US-08-922-774-42
3	316	67.1	740	3	US-09-632-711-42
4	316	67.1	740	3	US-09-632-703B-42
5	316	67.1	740	3	US-09-632-702-42
6	316	67.1	740	4	US-09-399-003-42
7	96.4	20.5	129	3	US-08-822-774-40
8	96.4	20.5	129	3	US-09-632-711-40
9	96.4	20.5	129	3	US-09-632-703B-40
10	96.4	20.5	129	3	US-09-632-702-40
11	96.4	20.5	129	4	US-09-399-003-40
12	82	17.4	84	3	US-08-822-774-36
13	82	17.4	84	3	US-09-632-711-36
14	82	17.4	84	3	US-09-632-703B-36
15	82	17.4	84	3	US-09-632-702-36
16	82	17.4	84	4	US-09-399-003-36
17	54.2	11.5	7218	1	US-08-232-463-14
18	38.6	8.2	579	4	US-09-540-236-1761
19	38.6	8.2	89047	4	US-09-596-002-34
20	37.2	7.9	561	4	US-09-248-796A-1970
21	37.2	7.9	846	4	US-09-248-796A-1971
22	35.2	7.5	267	4	US-09-248-796A-11979
23	34.4	7.3	601	4	US-09-949-016-204563
24	34.4	7.3	601	4	US-09-949-016-204564
25	34.4	7.3	601	4	US-09-949-016-204565
26	34.4	7.3	87734	4	US-09-949-016-17521
27	33.6	7.1	405	3	US-09-199-637A-242

28	33.6	7.1	765	3	US-09-199-637A-240	Sequence 240, App
29	33.6	7.1	3844	3	US-09-734-675-3	Sequence 3, Appl
30	33.6	7.1	42235	3	US-09-199-637A-1	Sequence 1, Appl
31	32.6	6.9	601	4	US-09-949-016-26855	Sequence 26855, A
32	32.6	6.9	601	4	US-09-949-016-55850	Sequence 55850, A
33	32.6	6.9	227390	4	US-09-949-016-12201	Sequence 12201, A
34	32.6	6.9	227391	4	US-09-949-016-13365	Sequence 13365, A
35	32.2	6.8	616	4	US-09-573-080A-417	Sequence 417, App
36	32.2	6.8	25656	4	US-09-949-016-13022	Sequence 13022, A
37	32	6.8	594	4	US-09-328-352-539	Sequence 539, App
38	32	6.8	23781	4	US-09-949-016-13446	Sequence 13446, A
39	32	6.8	39601	4	US-09-949-016-16045	Sequence 16045, A
40	32	6.8	147382	4	US-09-949-016-14624	Sequence 14624, A
41	31.8	6.8	491	3	US-08-896-164-11	Sequence 11, Appl
42	31.6	6.7	399	4	US-09-621-976-8976	Sequence 8976, Ap
43	31.6	6.7	609	4	US-09-252-991A-1091	Sequence 1091, Ap
44	31.6	6.7	3101	3	US-09-221-017B-87	Sequence 87, Appl
45	31.6	6.7	187136	4	US-09-949-016-17231	Sequence 17231, A

ALIGNMENTS

RESULT 1

US-09-399-003-70

; Sequence 70, Application US/09399003

; Patent No. 6734293

; GENERAL INFORMATION:

; APPLICANT: Hogrefe, Holly

; APPLICANT: Hansen, Connie J

; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes

; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying

; FILE REFERENCE: 4121.0116-02

; CURRENT APPLICATION NUMBER: US/09/399,003

; CURRENT FILING DATE: 1999-09-20

; PRIOR APPLICATION NUMBER: PCT/ US98/05497

; PRIOR FILING DATE: 1998-03-20

; PRIOR APPLICATION NUMBER: US 08/957,709

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: US 08/822,774

; PRIOR FILING DATE: 1997-03-21

; NUMBER OF SEQ ID NOS: 89

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 70

; LENGTH: 471

; TYPE: DNA

; ORGANISM: Pyrococcus furiosus

US-09-399-003-70

Query Match 100.0%; Score 471; DB 4; Length 471;

Best Local Similarity 100.0%; Pred. No. 2.7e-149;

Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCTACTTCCAGACTGGAAATACAGAAAGAAATACCTATAGAGCCATTTCTGAGAA	60
Db	1	ATGCTACTTCCAGACTGGAAATACAGAAAGAAATACCTATAGAGCCATTTCTGAGAA	60
QY	61	TCGCTCCAAACACAGAGGTTATGACCTCAGAGTGGGAGAGAGGCTTTGTTAAGGGGAAA	120
Db	61	TCGCTCCAAACACAGAGGTTATGACCTCAGAGTGGGAGAGAGGCTTTGTTAAGGGGAAA	120
QY	121	TTAATCGAGCTGGAAAGGAAAGTGGTTATTCCTCAAGGGAATACGCTTAAATC	180
Db	121	TTAATCGAGCTGGAAAGGAAAGTGGTTATTCCTCAAGGGAATACGCTTAAATC	180
QY	181	CTAACCTCGAGAGGATAAGTTGCCGAGATGTTATGGGGATATGAGATAGGAGC	240
Db	181	CTAACCTCGAGAGGATAAGTTGCCGAGATGTTATGGGGATATGAGATAGGAGC	240
QY	241	AGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGGTTGACCCAGGATGGGATGGA	300
Db	241	AGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGGTTGACCCAGGATGGGATGGA	300

301	AACTTAACACTAATGCTCTTACAATGCCCTCAAACTGTCGAATTAAGATATGAGAG	360	Qy
301	AACTTTAACACTAATGCTCTTACAATGCCCTCAAACTGTCGAATTAAGATATGAGAG	360	Db
361	AGATTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCCGGCAGAAACCCTTACAGAGGA	420	Qy
361	AGATTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCCGGCAGAAACCCTTACAGAGGA	420	Db
421	AACTATCAGGGGAGCACAAAGTTGATCGGTTTTCAAAGAGAAAGAACTCTAG	471	Qy
421	AACTATCAGGGGAGCACAAAGTTGATCGGTTTTCAAAGAGAAAGAACTCTAG	471	Db

```

RESULT 2
US-08-822-774-42
; Sequence 42, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CROSS-REFERENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-822-774-42

```

Qy	336	ACCTGTCGAATTAA	GATATGGAGAGAGATTTGTCAGATCCGATTTTATAAGGCTAGAGGG	395
Db	181	ACCTGTCGAATTAA	GATATGGAGAGAGATTTGTCAGATCCGATTTTATAAGGCTAGAGGG	240
Qy	396	TCCGGCAAGNAACCC	TTCACAGAGAAACTATCAGGGGAGCAACAAGTTAGCGTTTTTCAAA	455
Db	241	TCCGGCAAGNAACCC	TTCACAGAGAAACTATCAGGGGAGCAACAAGTTAGCGTTTTTCAAA	300
Qy	456	GAGAAAGAAACTCT	AGCTAG	471
Db	301	GAGAAAGAAACTCT	AGCTAG	316

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	67.1%	Score 316	DB 3	Length 740	
	Best Local Similarity	100.0%	Pred. No. 1.2e-96			
	Matches 316	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	156	TCCTCCAAAGGGATACGCCCTTAATCCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT	215			
Db		TCCTCCAAAGGGATACGCCCTTAATCCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT				
QY	216	TATGGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAGGGGTTATTTGGTCTCTTTTGC	275			
Db	61	TATGGGGGATATGAAGATNAGAGCAGTTTAGCAAGAGAGGGGTTATTTGGTCTCTTTTGC	120			
QY	276	TTGGGTTGACCCAGGATGGGATGGAACTTAACTAATGCTCTTCAATGCCTCAATGA	335			
Db	121	TTGGGTTGACCCAGATGGGATGAACTTAACTAATGCTCTTCAATGCCTCAATGA	180			

QY 276 TTGGTTGACCCAGGATGGATGAAACTTAACACATAATGCTCTACAAATGCTCAAAATGA 335
Db 121 TTGGTTGACCCAGGATGGATGAAACTTAACACATAATGCTCTACAAATGCTCAAAATGA 180
QY 336 ACCTGTGCAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 395
Db 181 ACCTGTGCAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 240
QY 396 TCCGGCAAGAACCCCTTACAGAGAACTATACAGGGAGCACAAGGTTAGCGTTTTCAAA 455
Db 241 TCCGGCAAGAACCCCTTACAGAGAACTATACAGGGAGCACAAGGTTAGCGTTTTCAAA 300
QY 456 GAGAAAGAACTCTAG 471
Db 301 GAGAAAGAACTCTAG 316
RESULT 4
US-09-632-703B-42
; Sequence 42, Application US/09632703B
; Patent No. 6379553
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF
; Protein Complexes, Isolated PEF Proteins, and Methods for P
; Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 24-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-703B-42
Query Match 67.1%; Score 316; DB 3; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCCAAGGAATAGCGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCGACCATGT 215
Db 1 TCCTCCAAGGAATAGCGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCGACCATGT 60
QY 216 TATGGGGGATATGAAGATAAGGAGCATTTAGCAAGAGAGGGGTTATTGGTTCTTTTTC 275

Db 61 TATGGGGGATATGAAGATAAGGAGCATTTAGCAAGAGAGGGGTTATTGGTTCTTTTTC 120
QY 276 TTGGTTGACCCAGGATGGATGAAACTTAACACATAATGCTCTACAAATGCTCAAAATGA 335
Db 121 TTGGTTGACCCAGGATGGATGAAACTTAACACATAATGCTCTACAAATGCTCAAAATGA 180
QY 336 ACCTGTGCAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 395
Db 181 ACCTGTGCAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 240
QY 396 TCCGGCAAGAACCCCTTACAGAGAACTATACAGGGAGCACAAGGTTAGCGTTTTCAAA 455
Db 241 TCCGGCAAGAACCCCTTACAGAGAACTATACAGGGAGCACAAGGTTAGCGTTTTCAAA 300
QY 456 GAGAAAGAACTCTAG 471
Db 301 GAGAAAGAACTCTAG 316
RESULT 5
US-09-632-702-42
; Sequence 42, Application US/09632702
; Patent No. 6444428
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,702
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-702-42
Query Match 67.1%; Score 316; DB 3; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCCAAGGAATAGCGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCGACCATGT 215

Db 1 TCCTCAAGGAATACGCTTAATCTAACCCTCGAGAGGATAAAGTTGCCCGACGATGT 60
QY 216 TATGGGGGATATCAAGATAAGGAGCAGTTTAGCAAGAGAGGGTTATTGGTCTTTTGC 275
Db 61 TATGGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAGGGTTATTGGTCTTTTGC 120
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAAATGCTCTACAATGCCTCAATGA 335
Db 121 TTGGTTGACCCAGGATGGGATGGAACCTTAACACTAAATGCTCTACAATGCCTCAATGA 180
QY 336 ACCTGTGCAATTAAAGATATGGAGAGAGATTGTGTCAGATCGCATTTTATAAGGCTAGAGGG 395
Db 181 ACCTGTGCAATTAAAGATATGGAGAGAGATTGTGTCAGATCGCATTTTATAAGGCTAGAGGG 240
QY 396 TCCGGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAGTTAGCGTTTTCAAA 455
Db 241 TCCGGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAGTTAGCGTTTTCAAA 300
QY 456 GAGAAAGAACTCTAG 471
Db 301 GAGAAAGAACTCTAG 316

RESULT 6
US-09-399-003-42
; Sequence 42, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hogrefe, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(740)
; OTHER INFORMATION: "n" is a, t, g, or c
US-09-399-003-42

Query Match 67.1%; Score 316; DB 4; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCAAGGGAATACGCTTAATCTAACCCTCGAGAGGATAAAGTTGCCCGACGATGT 215
Db 1 TCCTCAAGGGAATACGCTTAATCTAACCCTCGAGAGGATAAAGTTGCCCGACGATGT 60
QY 216 TATGGGGGATATCAAGATAAGGAGCAGTTTAGCAAGAGAGGGTTATTGGTCTTTTGC 275
Db 61 TATGGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAGGGTTATTGGTCTTTTGC 120
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAAATGCTCTACAATGCCTCAATGA 335
Db 121 TTGGTTGACCCAGGATGGGATGGAACCTTAACACTAAATGCTCTACAATGCCTCAATGA 180
QY 336 ACCTGTGCAATTAAAGATATGGAGAGAGATTGTGTCAGATCGCATTTTATAAGGCTAGAGGG 395
Db 181 ACCTGTGCAATTAAAGATATGGAGAGAGATTGTGTCAGATCGCATTTTATAAGGCTAGAGGG 240

QY 396 TCCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAGTTAGCGTTTTCAAA 455
Db 241 TCCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAGTTAGCGTTTTCAAA 300
QY 456 GAGAAAGAACTCTAG 471
Db 301 GAGAAAGAACTCTAG 316

RESULT 7
US-08-822-774-40/c
; Sequence 40, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREPE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-822-774-40

Query Match 20.5%; Score 96.4; DB 3; Length 129;
Best Local Similarity 99.0%; Pred. No. 1.2e-22;
Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAGAA 60
Db 98 ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAGAA 39
QY 61 TCGCTCCACACGAGGTTATGACCTCAGAGTGGGCAG 98
Db 38 TGGCTCCACACGAGGTTATGACCTCAGAGTGGGCAG 1
RESULT 8
US-09-632-711-40/c
; Sequence 40, Application US/09632711
; Patent No. 633165
; GENERAL INFORMATION:
; APPLICANT: HOGREPE, Holly

; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
 ; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
 ; and Methods for Purifying and Identifying Same
 ;
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
 ; STREET: 1200 G Street, N.W. Suite 700
 ; CITY: Washington
 ; STATE: D.C.
 ; ZIP: 20005
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/632,711
 ; FILING DATE: 04-Aug-2000
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/822,774
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KULIK, David J.
 ; REGISTRATION NUMBER: 36,576
 ; REFERENCE/DOCKET NUMBER: 1486/43163
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-8800
 ; TELEFAX: (202) 628-8844
 ;
 ; INFORMATION FOR SEQ ID NO: 40:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 129 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: DNA (Genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
 ;
 ; US-09-632-711-40
 ;
 ; Query Match 20.5%; Score 96.4; DB 3; Length 129;
 ; Best Local Similarity 99.0%; Pred. No. 1.2e-22;
 ; Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 ;
 ; QY 1 ATGCTACTCCAGACTCGAAATCAGAAAAGAAATCTATAGAGCCATTTCTGAAGAA 60
 ; Db 98 ATGCTACTCCAGACTCGAAATCAGAAAAGAAATCTATAGAGCCATTTCTGAAGAA 39
 ; QY 61 TCGCTCAACAGCAGGTTATGACCTCAGAGTGGGCAG 98
 ; Db 38 TGGCTCAACAGCAGGTTATGACCTCAGAGTGGGCAG 1
 ;
 ; RESULT 9
 ; US-09-632-703B-40/c
 ; Sequence 40, Application US/09632703B
 ; Patent No. 6379553
 ; GENERAL INFORMATION:
 ; APPLICANT: HOGREFE, Holly
 ; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF
 ; Protein Complexes, Isolated PEF Proteins, and Methods
 ; Identifying Same
 ;
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
 ; STREET: 1300 I Street, N.W.
 ; CITY: Washington
 ; STATE: D.C.
 ; ZIP: 20005-3315
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

```
;
;
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-09-632-702-40

Query Match 20.5%; Score 96.4; DB 3; Length 129;
Best Local Similarity 99.0%; Pred. No. 1.2e-22;
Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAA 60
Db 98 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAA 39

QY 61 TCGCTCCACCAGCAGGTTATGACCTCAGAGTGGCGAG 98
Db 38 TCGCTCCACCAGCAGGTTATGACCTCAGAGTGGCGAG 1

RESULT 11
US-09-399-003-40/c
; Sequence 40, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hogrefe, Holly
; APPLICANT: Hansen, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; FILE REFERENCE: 4121.0116-02
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/ US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 129
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
US-09-399-003-40

Query Match 20.5%; Score 96.4; DB 4; Length 129;
Best Local Similarity 99.0%; Pred. No. 1.2e-22;
Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAA 60
Db 98 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAA 39

QY 61 TCGCTCCACCAGCAGGTTATGACCTCAGAGTGGCGAG 98
Db 38 TCGCTCCACCAGCAGGTTATGACCTCAGAGTGGCGAG 1

RESULT 12
US-08-822-774-36
; Sequence 36, Application US/08822774
; Patent No. 6183997
```

```
;
;
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-822-774-36

Query Match 17.4%; Score 82; DB 3; Length 84;
Best Local Similarity 97.6%; Pred. No. 7.3e-18;
Matches 82; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 87 CAGAGTGGCGCAGAGGCTTTTGAAGGGGAAATTAATCGACGTGGAAAGGAGAAA 146
Db 1 CAGAGTGGCGCAGAGGCTTTTGAAGGGGAAATTAATCGACGTGGAAAGGAGAAA 60

QY 147 AGTCGTTATTCTCCCAAGGGAATA 170
Db 61 AGTCGTTATTCTCCCAAGGGAATA 84

RESULT 13
US-09-632-711-36
; Sequence 36, Application US/09632711
; Patent No. 6333165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,711
FILING DATE: 04-AUG-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/822,774
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 36:

US-09-632-711-36

Query Match 17.4%; Score 82; DB 3; Length 84;
Best Local Similarity 97.6%; Pred. No. 7.3e-18;
Matches 82; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 87 CAGAGTGGGCAGAGAGCGCTTTTGAAGGGGAAATTAATCGACGTGAAAAGGAAGAAA 146
Db 1 CAGAGTGGGCAGAGAGCGCTTTTGAAGGGGAAATTAATCGACGTGAAAAGGAAGAAA 60
QY 147 AGTCGTATTCTCCCAAGGGAATA 170
Db 61 AGTCGNTATTCTCCCAAGGGAATA 84

RESULT 14

US-09-632-703B-36
Sequence 36, Application US/09632703B
Patent No. 6379553
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: <Unknown>
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,703B
FILING DATE: 24-AUG-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/822,774
FILING DATE: 21-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: BARKER, M. Paul
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 4121.0116-04

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 849-6613
TELEFAX: (650) 849-6666
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-632-703B-36

Query Match 17.4%; Score 82; DB 3; Length 84;
Best Local Similarity 97.6%; Pred. No. 7.3e-18;
Matches 82; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 87 CAGAGTGGGCAGAGAGCGCTTTTGAAGGGGAAATTAATCGACGTGAAAAGGAAGAAA 146
Db 1 CAGAGTGGGCAGAGAGCGCTTTTGAAGGGGAAATTAATCGACGTGAAAAGGAAGAAA 60
QY 147 AGTCGTATTCTCCCAAGGGAATA 170
Db 61 AGTCGNTATTCTCCCAAGGGAATA 84

RESULT 15

US-09-632-702-36
Sequence 36, Application US/09632702
Patent No. 6444428
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,702
FILING DATE: 04-AUG-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-632-702-36
Query Match      17.4%; Score 82; DB 3; Length 84;
Best Local Similarity 97.6%; Pred. No. 7.3e-18;
Matches 82; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      87 CAGAGTGGGCAGAGAGGCTTTGTTAAGGGGAAATTAATCGACGTGGAAGGAAAGGAAA 146
Db      1 CAGAGTGGGCAGAGAGGCTTTGTTAAGGGGAAATTAATCGACGTGGAAGGAAAGGAAA 60

QY      147 AGTCGTTATTCTCCCAAGGGAATA 170
Db      61 AGTCGNTATTCTCCCAAGGGAATA 84
```

Search completed: July 2, 2005, 06:29:18
Job time : 147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 14:45:07 ; Search time 420 Seconds
(without alignments)

6638.567 Million cell updates/sec

Title: US-08-957-709A-70

Perfect score: 471

Sequence: 1 atgctactccagactggaa.....caaagagaagaactcttag 471

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_16Dec04:*
- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	471	100.0	471	2	AAV63860
2	280.8	59.6	265118	5	AAH41227
3	256.2	54.4	110000	12	ADN46845_12
4	256.2	54.4	110000	12	ADN47591_08
5	256.2	54.4	110000	12	ADN46123_12
6	256.2	54.4	110000	12	ADN47209_08
7	256.2	54.4	110000	12	ADN46464_12
8	256.2	54.4	110000	12	ADN47960_08
9	49	10.4	419	6	ABX66923
10	44.8	9.5	3382	2	AAQ62300
11	44.2	9.4	2006	2	AAV90555
12	44.2	9.4	2588	2	AAV90871
13	40.2	8.5	2964	13	ADR89139
14	39.2	8.3	597	12	ADK16664
15	39.2	8.3	110000	12	ADK16049_2
16	38.6	8.2	579	12	ADJ04075
17	38.6	8.2	89047	4	AAF28547
18	37.4	7.9	9202	1	AAV70608
19	37.2	7.9	110021	13	ABD32859
20	35.4	7.5	1097	6	ABQ91119

C	21	35	7.4	3257	2	AAT14352	Rat14352 Nk2 rece
C	22	35	7.4	3257	2	AAV38546	Aav38546 Mouse rec
	23	34.8	7.4	570	4	AAI17857	Aai17857 Probe #77
	24	34.8	7.4	570	4	ABA61043	Aba61043 Human foe
	25	34.8	7.4	570	4	ABA62813	Aba62813 Human foe
	26	34.8	7.4	570	4	AAI42828	Aai42828 Probe #11
	27	34.8	7.4	570	4	AAI40941	Aai40941 Probe #96
	28	34.8	7.4	570	4	ABA28956	Aba28956 Probe #74
	29	34.8	7.4	570	4	ABA30099	Aba30099 Probe #85
	30	34.8	7.4	570	4	AAK35225	Aak35225 Human bon
	31	34.8	7.4	570	4	AAK09336	Aak09336 Human bra
	32	34.8	7.4	570	4	ABS34971	Ab34971 Human liv
C	33	34.6	7.3	90435	12	ADQ59524	Adq59524 Human can
	34	34	7.2	110000	6	ABX08336_11	Continuation (12 o
	35	34	7.2	110000	12	ADJ25985_11	Continuation (12 o
	36	34	7.2	110000	12	ADN97989_11	Continuation (12 o
	37	34	7.2	110000	12	ADO50281_11	Continuation (12 o
C	38	33.8	7.2	349980	6	ABQ81849	Abq81849 Bifidobac
C	39	33.6	7.1	405	2	AAK98102	Aax98102 Nucleotid
	40	33.6	7.1	764	6	ABQ91833	Abq91833 M. capsul
	41	33.6	7.1	765	2	AAK98101	Aax98101 Nucleotid
	42	33.6	7.1	12269	4	AAK46472	Aas46472 Tumour su
	43	33.6	7.1	12269	6	ABL33411	Ab133411 Human imm
C	44	33.6	7.1	18650	4	AAK79108	Aak79108 Human imm
C	45	33.6	7.1	18671	4	AAK90763	Aak90763 Human dig

ALIGNMENTS

RESULT 1
AAV63860
ID AAV63860 standard; DNA; 471 BP.
AC AAV63860;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 (dUTPase) component DNA.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication; ss.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
(STRA-) STRATAGENE.
Hogrefe H, Hansen CJ;
WPI; 1998-542284/46.
P-PSDB; AAW72847.
Polymerase enhancing factor proteins, extracts and complexes - improve
the polymerisation activity of nucleic acid polymerases, for use in
amplification, sequencing and replication.
Claim 17; Page 43; 161pp; English.
This DNA sequence encodes the P45 component (see AAW72847) of the
polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. PCR
primers (see AAV63861-64) based on an N-terminal peptide (see AAW72846)
of P45 were used to amplify P. furiosus genomic DNA, and sequencing was
carried out on purified PCR products and plasmid mini-preps. P45 and P40

(see AAW72844) are the predominant components of PBF, which acts to enhance the activity of P. furiosus DNA polymerase, thereby providing replication products of greater length and purity. P45 was identified as a dUTPase, and possesses polymerase enhancing activity. The invention provides novel extracts, proteins and complexes that improve the polymerisation activity of nucleic acid polymerases, as well as DNA constructs and antibodies. Also included are methods for identifying compositions with polymerase enhancing activity, for purifying and using these compositions, and specific extracts, proteins and complexes that function to enhance polymerase activity. Nucleic acid polymerase reactions can be enhanced (claimed) by mixing a nucleic acid template, at least 1 polymerase and a composition having polymerase enhancing activity. Kits are provided for replicating nucleic acids. The kits can be used in site-directed mutagenesis, nucleic acid sequencing or amplification (preferably PCR or RT-PCR). Isolated DNA can be used in the recombinant production of P45. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 471 BP; 156 A; 80 C; 124 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 471; DB 2; Length 471;
 Best Local Similarity 100.0%; Pred. No. 4.5e-142;
 Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60
 DB 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60

QY 61 TCGCTCCAAACCGACGGTTATGACCTCAGAGTGGGCGAGAGAGGCTTTTGAAGGGGAAA 120
 DB 61 TCGCTCCAAACCGACGGTTATGACCTCAGAGTGGGCGAGAGAGGCTTTTGAAGGGGAAA 120

QY 121 TTAATCGAGTGGAAAGGAAAGAAAGTGGTTATTCCTCCAAAGGAATACGCTTTAATC 180
 DB 121 TTAATCGAGTGGAAAGGAAAGTGGTTATTCCTCCAAAGGAATACGCTTTAATC 180

QY 181 CTAACCTCGAGAGGATAAGTTGCCCGACGATGTTATGGGGATATGAAGATAAGGAGC 240
 DB 181 CTAACCTCGAGAGGATAAGTTGCCCGACGATGTTATGGGGATATGAAGATAAGGAGC 240

QY 241 AGTTTACGAGGAGGAGGTTATGTTCTTTTTCCTTGGGTTGACCCAGGATGGGATGA 300
 DB 241 AGTTTACGAGGAGGAGGTTATGTTCTTTTTCCTTGGGTTGACCCAGGATGGGATGA 300

QY 301 AACTTAACTAATGCTCTACAAATGCTCAAAATGAACTGTGCAATTAAGATATGAGAG 360
 DB 301 AACTTAACTAATGCTCTACAAATGCTCAAAATGAACTGTGCAATTAAGATATGAGAG 360

QY 361 AGATTTGTCGAGATCGCATTTAAGGCTAGAGGGTCCGCGAAGAACCCCTTACAGAGGA 420
 DB 361 AGATTTGTCGAGATCGCATTTAAGGCTAGAGGGTCCGCGAAGAACCCCTTACAGAGGA 420

QY 421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCAAAGAGAAAGAACTCTAG 471
 DB 421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCAAAGAGAAAGAACTCTAG 471

RESULT 2

AAH41227

ID AAH41227 standard; DNA; 265118 BP.

XX

AC AAH41227;

XX

DT 29-OCT-2001 (first entry)

XX

DE Pyrococcus abyssi genomic fragment #6.

XX

KW Hyperthermophilic archaeon; hyperthermophilic protein; ds.

XX

OS Pyrococcus abyssi.

XX

FH Key Location/Qualifiers

FT misc_feature 1..49980

FT /*tag= a
 FT /note= "This sequence overlaps with the 3' end of
 FT AAH41226"
 XX
 PN FR2792651-Al.
 XX
 PD 27-OCT-2000.
 XX
 PF 21-APR-1999; 99FR-00005034.
 XX
 PR 21-APR-1999; 99FR-00005034.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (IFRE-) IFREMER INST FR RECH EXPL MER.
 XX
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Heilig R;
 XX
 DR WPI; 2001-126236/14.
 XX
 PT New nucleotide sequences isolated from Pyrococcus abyssi encode proteins
 PT useful in industry.
 XX
 PS Claim 1; Page 593-665; 1657pp; French.

XX The present invention relates to the genomic sequence of Pyrococcus
 CC abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
 CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
 CC vents. The present sequence is a fragment of the genomic sequence of P.
 CC abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41226.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade. Note: This patent is in the same patent family as
 CC WO200005062, which contains additional sequences as shown in AAB99132-
 CC AAB99143, AAH75903-AAH75920 and AAG66436
 XX
 SQ Sequence 265118 BP; 75226 A; 61302 C; 54005 G; 74585 T; 0 U; 0 Other;

Query Match 59.6%; Score 280.8; DB 5; Length 265118;
 Best Local Similarity 75.0%; Pred. No. 3.8e-79;
 Matches 351; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60
 DB 262045 ATGCTCTCTCCAGACTGGAAATAGGAAGAGATTTTAATAAGCCATTTCTCAGAGAA 262104

QY 61 TCGCTCCAAACCGACGAGTTATGACCTCAGAGTGGGCGAGAGAGGCTTTTGTAAAGGGGAAA 120
 DB 262105 TCGCTTCAACGAGCTGGCTACGACTTAAGGGTAGGCAAGGAAGCTTACATTCAGGGAAAG 262164

QY 121 TTAATCGAGCTGGAAAGGAAAGTGGTTATTCCTCCAAAGGAATACGCTTTAATC 180
 DB 262165 TTAATCGAGCTGGAAAGGAAAGTGGTTATTCCTCCAAAGGAATACGCTTTAATC 180

QY 181 CTAACCTCGAGAGGATAAGTTGCCCGACGATGTTATGGGGATATGAAGATAAGGAGC 240
 DB 262225 CTAACCTCGAGAGGATAAGCTTCCAGATGATATTATGGGAGACATGAAGATAAGGAGC 262284

QY 241 AGTTTACGAGGAGGAGGTTATGTTCTTTTTCCTTGGGTTGACCCAGGATGGGATGA 300
 DB 262285 AGTCTAGCTAGGAAAGGTTCTAGGCTCTTTTGTGATGGGTAGACCCGGGACGGT 262344

QY 301 AACTTAACTAATGCTCTACAAATGCTCAAAATGAACTGTGCAATTAAGATATGAGAG 360
 DB 262345 AACTTAACTTGTATGCTTTTACAAATGCAATCGGAAAGGAGGTAATTTTAAGGTACAAAGAG 262404

QY 361 AGATTTGTCGAGATCGCATTTAAGGCTAGAGGGTCCGCGAAGAACCCCTTACAGAGGA 420
 DB 262405 AGGTTTGTCCAGATAGCCTTCTTCCAGGCTTTAGGCTCCCGCGAAGAAATCCATACAGGGGC 262464

QY 421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCAAAGAGAAAGAACTC 468
 DB 262465 AACTATCAGGAAAGCAGGAGAAATAGTCTCTCTCAAGAGATCTTAACTC 262512

RESULT 3

ADN46845_12/c
Continuation (13 of 21) of ADN46845 from base 1200001 (Thermococcus kodakaraensis KOD1 gen
WP Sequence split into 21 fragments LOCUS ADN46845 Accession Adn46845

WP	Fragment Name	Begin	End
WP	ADN46845_00	1	110000
WP	ADN46845_01	100001	210000
WP	ADN46845_02	200001	310000
WP	ADN46845_03	300001	410000
WP	ADN46845_04	400001	510000
WP	ADN46845_05	500001	610000
WP	ADN46845_06	600001	710000
WP	ADN46845_07	700001	810000
WP	ADN46845_08	800001	910000
WP	ADN46845_09	900001	1010000
WP	ADN46845_10	1000001	1110000
WP	ADN46845_11	1100001	1210000
WP	ADN46845_12	1200001	1310000
WP	ADN46845_13	1300001	1410000
WP	ADN46845_14	1400001	1510000
WP	ADN46845_15	1500001	1610000
WP	ADN46845_16	1600001	1710000
WP	ADN46845_17	1700001	1810000
WP	ADN46845_18	1800001	1910000
WP	ADN46845_19	1900001	2010000
WP	ADN46845_20	2000001	2089378

Query Match 54.4%; Score 256.2; DB 12; Length 110000;

Best Local Similarity 72.2%; Pred. No. 2.4e-71;

Matches 333; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY	1	ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACCTATATAGAGCAATTTCTGAAGAA	60
DB	40800	ATGATGCTTCCGATTTGGAAGATTAGNAGGAATTTTGTATCGAGCTTTTCAGTGAAGAA	40741
QY	61	TCGCTCCAAACAGCAGGTTATGACCTCAGAGTCGGCAGAGAGGCTTTTGTAAAGGGAAA	120
DB	40740	TCCTCGAGCTCGAGGATATGACCTAAGGGTTGGAGGCGAAGCTTACGTAAATGGTAAA	40681
QY	121	TTAATCGAGTGAAGAGGAAAGTTCCTTCCCAAGGAATACGCTTAATC	180
DB	40680	ATACTCGAGCTTAAGGATTCCGGGGAGTTACAATACCCCAAGACTTACGCTTTGGTT	40621
QY	181	CTAACCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGATATGAAGATAAGGAGC	240
DB	40620	TTGACTCTCGAGAGGATCAAGCTTCCGGATGATGTAATGGCGACATGAAGCTTAGGAGC	40561
QY	241	AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGTTGACCCAGGATGGGATGGA	300
DB	40560	AGTCTAGCCAGAGAGGGCTTAATCGGTTCTTTTGGCTGGGTTGACCTGGATGGATGGA	40501
QY	301	AACCTTAACACTAATGCTCTCAATAGCTCAATGAACCTGTGCAATTAAGATATGAGAG	360
DB	40500	AATCTCACTCTAGCCCTGTTCAACGCTTCCACGAGTCTGTGGAACCTCAAAATACGGCGAG	40441
QY	361	AGATTTGTGCAGATCGCATTTATAAGCTAGAGGTCGGCAAGAAACCTTACAGAGGA	420
DB	40440	CGCTTTGTTTCAGATAGGTTTATCCGGCTGGAGGGCTGCTAAAAACCCCTATCGTGA	40381
QY	421	AACCTATCAGGGAGCACAGGTTTAGCGTTTTCAAAGAGAAA	461
DB	40380	AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAAGGAA	40340

RESULT 4

ADN47591_08
Continuation (9 of 21) of ADN47591 from base 800001 (Thermococcus kodakaraensis KOD1 gen
WP Sequence split into 21 fragments LOCUS ADN47591 Accession Adn47591

WP	Fragment Name	Begin	End
WP	ADN47591_00	1	110000
WP	ADN47591_01	100001	210000

WP	ADN47591_02	200001	310000
WP	ADN47591_03	300001	410000
WP	ADN47591_04	400001	510000
WP	ADN47591_05	500001	610000
WP	ADN47591_06	600001	710000
WP	ADN47591_07	700001	810000
WP	ADN47591_08	800001	910000
WP	ADN47591_09	900001	1010000
WP	ADN47591_10	1000001	1110000
WP	ADN47591_11	1100001	1210000
WP	ADN47591_12	1200001	1310000
WP	ADN47591_13	1300001	1410000
WP	ADN47591_14	1400001	1510000
WP	ADN47591_15	1500001	1610000
WP	ADN47591_16	1600001	1710000
WP	ADN47591_17	1700001	1810000
WP	ADN47591_18	1800001	1910000
WP	ADN47591_19	1900001	2010000
WP	ADN47591_20	2000001	2089378

Query Match 54.4%; Score 256.2; DB 12; Length 110000;

Best Local Similarity 72.2%; Pred. No. 2.4e-71;

Matches 333; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY	1	ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACCTATATAGAGCAATTTCTGAAGAA	60
DB	48578	ATGATGCTTCCGATTTGGAAGATTAGAAAGGAAATTTTGTATCGAGCTTTTCAGTGAAGAA	48637
QY	61	TCGCTCCAAACAGCAGGTTATGACCTCAGAGTCGGCAGAGAGGCTTTTGTAAAGGGAAA	120
DB	48638	TCCTCGAGCTCGAGGATATGACCTAAGGGTTGGAGGCGAAGCTTACGTAAATGGTAAA	48697
QY	121	TTAATCGAGTGAAGAGGAAAGTTCCTTCCCAAGGAATACGCTTAATC	180
DB	48698	ATACTCGAGCTTAAGGATTCCGGGGAGTTACAATACCCCAAGACTTACGCTTTGGTT	48757
QY	181	CTAACCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGATATGAAGATAAGGAGC	240
DB	48758	TTGACTCTCGAGAGGATCAAGCTTCCGGATGATGTAATGGCGACATGAAGCTTAGGAGC	48817
QY	241	AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGTTGACCCAGGATGGGATGGA	300
DB	48818	AGTCTAGCCAGAGAGGGCTTAATCGGTTCTTTTGGCTGGGTTGACCTGGATGGATGGA	48877
QY	301	AACCTTAACACTAATGCTCTCAATAGCTCAAAATGAACCTGTGCAATTAAGATATGGAGAG	360
DB	48878	AATCTCACTCTAGCCCTGTTCAACGCTTCCACGAGTCTGTGGAACCTCAAAATACGGCGAG	48937
QY	361	AGATTTGTGCAGATCGCATTTATAAGCTAGAGGTCGGCAAGAAACCCCTTACAGAGGA	420
DB	48938	CGCTTTGTTTCAGATAGGTTTATCCGGCTGGAGGGCTGCTAAAAAACCCCTATCGTGA	48997
QY	421	AACCTATCAGGGAGCACAGGTTTAGCGTTTTCAAAGAGAAA	461
DB	48998	AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAAGGAA	49038

RESULT 5

ADN46123_12/c
Continuation (13 of 21) of ADN46123 from base 1200001 (Thermococcus kodakaraensis KOD1
WP Sequence split into 21 fragments LOCUS ADN46123 Accession Adn46123

WP	Fragment Name	Begin	End
WP	ADN46123_00	1	110000
WP	ADN46123_01	100001	210000
WP	ADN46123_02	200001	310000
WP	ADN46123_03	300001	410000
WP	ADN46123_04	400001	510000
WP	ADN46123_05	500001	610000
WP	ADN46123_06	600001	710000
WP	ADN46123_07	700001	810000
WP	ADN46123_08	800001	910000
WP	ADN46123_09	900001	1010000
WP	ADN46123_10	1000001	1110000

	ADN46123_11	1100001	1210000	
WP	ADN46123_12	1200001	1310000	
WP	ADN46123_13	1300001	1410000	
WP	ADN46123_14	1400001	1510000	
WP	ADN46123_15	1500001	1610000	
WP	ADN46123_16	1600001	1710000	
WP	ADN46123_17	1700001	1810000	
WP	ADN46123_18	1800001	1910000	
WP	ADN46123_19	1900001	2010000	
WP	ADN46123_20	2000001	2089378	
	Query Match	54.4%;	Score 256.2; DB 12; Length 110000;	
	Best Local Similarity	72.2%;	Pred. No. 2.4e-71;	
	Matches 333; Conservative	0;	Mismatches 128; Indels 0; Gaps 0;	
Qy	1 ATGCTACTTCCAGACTGGAAATCAGAAAAGAATACTTTATAGAGCCATTTCCTCAAGAA	60		
Dd	40800 ATGATCGTCCCCGATTGGGAAGATAGAAAGGAATTTTGATCGAGCCTTCAGTGAAGA	40741		
Qy	61 TCGCTCCAACACAGCAGGTTCATGACCCTCAGAGTGGGCAGAGAGCCTTTTGTTAAGGGGAAA	120		
Dd	40740 TC CCTCAGCCTGCAGCATATGACCTAAGGGTTGGAGGCCAAGCTTAGCTAAATGCTAAA	40681		
Qy	121 TTAAATCGAGTGGAAAGAGGAAGAAGTCGTTATTCTCTCAAAGGAATACGCCCTTAATC	180		
Dd	40680 ATA CT CGAG GTT AAG GA N T C AG GG GG AG T T A CA AT ACC CC C AA A G AC T T AC GC CT TG GT T	40621		
Qy	181 CTAAACCTCGAGAGGATAAAGTTGCCCGACGATGTTATGGGGGATATGAAGATAAGGAGC	240		
Dd	40620 TTCACCTCTCGAGAGGATCAAGCTTCGGATGATGTATNTGGGCGCATGNAGCTTAGGAGC	40561		
Qy	241 AGTTTAGCAAGAGAGGGGTATTGTTTTTTTTTTGGTTGGATCCCAGGATGGGATGGA	300		
Dd	40560 AGCTAGCCAGAGAGGGCTTAATCGGTTCTTTTGCCTGGGTTGACCTCGGATGGGATGGA	40501		
Qy	301 AACCTAACACTAATGCTCTACAATGCCTCAATGAACCTGTGCAATTAAAGATATGAGAG	360		
Dd	40500 AATCTCACTCTAGCCCTGTTCACCGTTCCACGAGTCTGTGGAACTCAAAATACGGCGAG	40441		
Qy	361 AGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGA	420		
Dd	40440 CGCTTTGTTCAGATAGCGTTTATCCGGCTGGAGGGCGCTGTCAAAAAACCCTATCGTGA	40381		
Qy	421 AACTATCAGGGGAGCACAAAGTTTAGCGTTTTTCAAGAGAAA	461		
Dd	40380 AATTACAGGAAGCAAAACATCTAGCGCTCTCAAAAGGAA	40340		

RESULT 6			ADNA47209 08		Continuation (9 of 21) of ADNA47209 from base 800001 (Thermococcus kodakaraensis KOD1 ger	
WP	Sequence	split into 21 fragments	LOCUS	ADNA47209	Accession	ADNA47209
WP	Fragment Name	Begin	End			
WP	ADNA47209_00	1	110000			
WP	ADNA47209_01	100001	210000			
WP	ADNA47209_02	200001	310000			
WP	ADNA47209_03	300001	410000			
WP	ADNA47209_04	400001	510000			
WP	ADNA47209_05	500001	610000			
WP	ADNA47209_06	600001	710000			
WP	ADNA47209_07	700001	810000			
WP	ADNA47209_08	800001	910000			
WP	ADNA47209_09	900001	1010000			
WP	ADNA47209_10	1000001	1110000			
WP	ADNA47209_11	1100001	1210000			
WP	ADNA47209_12	1200001	1310000			
WP	ADNA47209_13	1300001	1410000			
WP	ADNA47209_14	1400001	1510000			
WP	ADNA47209_15	1500001	1610000			
WP	ADNA47209_16	1600001	1710000			
WP	ADNA47209_17	1700001	1810000			
WP	ADNA47209_18	1800001	1910000			
WP	ADNA47209_19	1900001	2010000			

WP	ADN47209_20	2000001	2089378	
	Query Match	54.4%	Score 256.2; DB 12; Length 110000;	
	Best Local Similarity	72.2%;	Pred. No. 2.4e-71;	
	Matches 333; Conservative	0; Mismatches 128; Indels	0; Gaps	0;
Qy	1	ATGCTACTTCGAGACTGGHAAATCAGAAAAGNAATACTTATAGAGCCATTTCCTCAAGAA	60	
Db	48578	ATGATCTCTCCGATTGGAGAGATTAGAAGGAAATTTTGTATCGAGCCTTCAGTGAAGAA	48637	
Qy	61	TGCTCTCAACACGAGAGTTATCACTCTCAGAGTGGCAGAGAGCGCTTTTGTGTTAAGGGGAAA	120	
Db	48638	TCCTCTGACCTGCGAGATATGACCTAAGGTTGGAGGCGAAGCTTACGTAAATGGTAAA	48697	
Qy	121	TTAATCGAGCTGAAAAGGAAGGAAGAAAGTCGTTATTCTCTCCAAGGGAATACGCCTTAATC	180	
Db	48698	ATATCTGAGCTTAAGGATTCAGGGGGAGTTACAATACCCCCAAAGACTTACGCCTTGGTT	48757	
Qy	181	CTAACCCCTCGAGAGGATAAAGTTGCCGCGAGATGTTATGGGGATATGAAAGATAAGGAGC	240	
Db	48758	TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAATGGGCGACATGAAGCTTAGGAGC	48817	
Qy	241	AGTTTAGCAAGAGAGGGGTAACTGGTCTCTTTTGTCTGGGTTGACCCAGGATGGGATCGA	300	
Db	48818	AGTCTAGCCAGAGAGGGCTTAATCGGTTCTTTTGTGCTGGTTGACCCCTGGATGGGATGGA	48877	
Qy	301	AACTTAACACTAATGCTCTACAATGCCTCAAAATGAACCTGTTCGAATTAAGATATGGAGAG	360	
Db	48878	AATCTCACTCTAGCCCTGTTCAACCGTTCACAGAGTCTGTGGAATCAAAATACGCCGAG	48937	
Qy	361	AGATTTGTGCAGATCGCATTTATAAGGCTAGAGGTCCGGCAAGAAACCCCTTACAGAGGA	420	
Db	48938	CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCCCTGCTTAAAAAACCCCTATCGTGA	48997	
Qy	421	AACTATCAGGGGAGCAAGGTTAGCGTTTTCAAAAGAGAA	461	
Db	48998	AATTACACAGGGAAGCAACATCTAGCGCTCTCAAAAGGAA	49038	

RESULT	7	ADN46464_12/c	Continuation (13 of 21) of ADN46464 from base 1200001 (Thermococcus kodakaraensis KOD1)
WP	Sequence split into 21 fragments	LOCUS ADN46464 Accession Adn46464	
WP	Fragment Name	Begin	End
WP	ADN46464_00	1	110000
WP	ADN46464_01	100001	210000
WP	ADN46464_02	200001	310000
WP	ADN46464_03	300001	410000
WP	ADN46464_04	400001	510000
WP	ADN46464_05	500001	610000
WP	ADN46464_06	600001	710000
WP	ADN46464_07	700001	810000
WP	ADN46464_08	800001	910000
WP	ADN46464_09	900001	1010000
WP	ADN46464_10	1000001	1110000
WP	ADN46464_11	1100001	1210000
WP	ADN46464_12	1200001	1310000
WP	ADN46464_13	1300001	1410000
WP	ADN46464_14	1400001	1510000
WP	ADN46464_15	1500001	1610000
WP	ADN46464_16	1600001	1710000
WP	ADN46464_17	1700001	1810000
WP	ADN46464_18	1800001	1910000
WP	ADN46464_19	1900001	2010000
WP	ADN46464_20	2000001	2089378
Query Match	54.4%	Score 256.2; DB 12;	Length 110000;
Best Local Similarity	72.2%;	Pred. No. 2.4e-71;	
Matches 333;	Conservative	0; Mismatches 128;	Indels 0; Gaps 0;
Qy	1	ATGCTACTTCAGACTCGAAATCCAGAATAAACAAGAAATACATTATAGAGGCCATTTTCTGAAGAA	60
Dd	40800	ATGATGTCTCCCAGATTGGAGATTAGAAAGAAAATTTTGATCGAGSCTTTCAGTGAAGAA	40741


```
QY 61 TCCTCCACCAGCAGGTTATGACCTCAGAGTGGGCGAGAGAGGCTTTTGTAAAGGGGAAA 120
DB 40740 TCCCTGCAGCCTGCAGGATATGACCTAAGGGTTGGAGGCGAAGCTTACGCTAAATGTTAAA 40681
QY 121 TTAATCGAGCTGGAAGAAGGAAAAGTGGTTATTTCTCCAAAGGGAATACGCTTTAATC 180
DB 40680 ATACTCGAGTTAAGGATTCAGGGGAGTTACAAATACCCCAAGACTTACGCTTGGTT 40621
QY 181 CTAACCTCCAGAGGATAAAGTTGCCGACGATGTTTATGGGGGATATGAAGATAAGGAGC 240
DB 40620 TTGACTCTCAGAGGATCAAGCTTCCGATGATGTAATGGGCGACATGAAGCTTACGAGC 40561
QY 241 AGTTTAGCAAGAGAGGGTTATGTTCTTTTTCCTTGGGTTGACCCAGAGATGGGATGGA 300
DB 40560 AGCTTAGCCAGAGAGGGCTTAATCGGTTCTTTTTCCTTGGGTTGACCCCTGGATGGATGGA 40501
QY 301 AACTTAACACTAATGCTCTACAAATGCTCAAAATGAACCTGTGCAATTAAGATATGGAG 360
DB 40500 AATCTCACTCTAGCCCTGTTCACAGCTTCCACGAGTCTGTGGAATCAAAATACGGGAG 40441
QY 361 AGATTTGTGCAGATCGCATTTATAGGCTAGAGGGTCCGCGAAGAAACCCCTTACAGAGGA 420
DB 40440 CGCTTTGTTTCAGATAGCGTTTATCCGCTCGAGGGGCTGCTAAAAACCCCTATCTGGA 40381
QY 421 AACTATCAGGGAGCAAGGTTAGCGTTTTCAAAGAGAAA 461
DB 40380 AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAAGGAA 40340
```

RESULT 8

ADN47960_08
Continuation (9 of 21) of ADN47960 from base 800001 (Thermococcus kodakaraensis KOD1 Gen

```
WP Sequence split into 21 fragments LOCUS ADN47960 Accession Adn47960
WP Fragment Name Begin End
WP ADN47960_00 1 110000
WP ADN47960_01 100001 210000
WP ADN47960_02 200001 310000
WP ADN47960_03 300001 410000
WP ADN47960_04 400001 510000
WP ADN47960_05 500001 610000
WP ADN47960_06 600001 710000
WP ADN47960_07 700001 810000
WP ADN47960_08 800001 910000
WP ADN47960_09 900001 1010000
WP ADN47960_10 1000001 1110000
WP ADN47960_11 1100001 1210000
WP ADN47960_12 1200001 1310000
WP ADN47960_13 1300001 1410000
WP ADN47960_14 1400001 1510000
WP ADN47960_15 1500001 1610000
WP ADN47960_16 1600001 1710000
WP ADN47960_17 1700001 1810000
WP ADN47960_18 1800001 1910000
WP ADN47960_19 1900001 2010000
WP ADN47960_20 2000001 2089378
```

```
Query Match 54.4%; Score 256.2; DB 12; Length 110000;
Best Local Similarity 72.2%; Pred. No. 2.4e-71;
Matches 333; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAATAATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAAGAA 60
DB 48578 ATGATGCTTCCCGATTTGGAGATTAGAAAAGGAAATTTTGATCGAGCTTTCAGTGAAGAA 48637
QY 61 TCCTCCACCAGCAGGTTATGACCTCAGAGTGGGCGAGAGAGGCTTTTGTAAAGGGGAAA 120
DB 48638 TCCTGCAGCCTGCAGGATATGACCTAAGGGTTGGAGGCGAAGCTTACGCTAAATGTTAAA 48697
QY 121 TTAATCGAGCTGGAAGAAGGAAAAGTGGTTATTTCTCCAAAGGGAATACGCTTTAATC 180
DB 48698 ATACTCGAGTTAAGGATTCAGGGGAGTTACAAATACCCCAAGACTTACGCTTGGTT 48757
```

```
QY 181 CTAACCTCCAGAGGATAAAGTTGCCGACGATGTTTATCGGGGATATGAAGATAAGGAGC 240
DB 48758 TTGACTCTCAGAGGATCAAGCTTCCGATGATGTAATGGGCGACATGAAGCTTACGAGC 48817
QY 241 AGTTTAGCAAGAGAGGGGTTATGTTCTTTTTCCTTGGGTTGACCCAGAGATGGGATGGA 300
DB 48818 AGCTTAGCCAGAGAGGGCTTAATCGGTTCTTTTTCCTTGGGTTGACCCCTGGATGGATGGA 48877
QY 301 AACTTAACTAATGCTCTACAAATGCTCAAAATGAACCTGTGCAATTAAGATATGGAGAG 360
DB 48878 AATCTCACTCTAGCCCTGTTCACAGCTTCCAAACGAGTCTGTGGAATCAAAATACGGCGAG 48937
QY 361 AGATTTGTGCAGATCGCATTTATAGGCTAGAGGGTCCGCGAAGAAACCCCTTACAGAGGA 420
DB 48938 CGCTTTGTTTCAGATAGCGTTTATCCGCTCGAGGGGCTGCTAAAAAACCCCTATCTGGA 48997
QY 421 AACTATCAGGGAGCAACAGGTTAGCGTTTTCAAAGAGAAA 461
DB 48998 AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAAGGNA 49038
```

RESULT 9

```
ADN66923
ID ADN66923 standard; DNA; 419 BP.
XX
AC ADN66923;
XX
DT 07-MAY-2003 (first entry)
XX
DE Helicobacter pylori selected interacting domain (SID) DNA #1522.
XX
KW Protein-protein interaction; ulcer; selected interacting domain; SID;
  gene; db.
XX
OS Helicobacter pylori.
XX
PN WO200266501-A2.
XX
PD 29-AUG-2002.
XX
PF 28-DEC-2001; 2001WO-EP015428.
XX
PR 02-JAN-2001; 2001US-0259302P.
XX
PA (HYBR-) HYBRIGENICS.
  (INSP ) INST PASTEUR.
XX
PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX
WPI; 2002-674910/72.
DR P-PSDB; ABU52179.
XX
PT New complexes of protein-protein interactions in Helicobacter pylori,
  useful for identifying modulating compounds for treating or preventing
  ulcers in mammals.
XX
PS Claim 7; Page 457; 642pp; English.
```

```
XX
CC The invention describes a complex of protein-protein interactions in
  Helicobacter pylori selected from 421 complexes given in the
  specification. The complex of protein-protein interactions are useful for
  screening for agents which modulate the interaction of proteins.
  CC Modulating compounds which binds to a targeted bacterial protein may be
  used for treating or preventing ulcers in a human or animal. This
  CC sequence encodes a selected interacting domain (SID), identified via
  protein-protein interactions
```

SQ Sequence 419 BP; 129 A; 73 C; 99 G; 118 T; 0 U; 0 Other;

```
Query Match 10.4%; Score 49; DB 6; Length 419;
Best Local Similarity 48.7%; Pred. No. 3.4e-05;
Matches 133; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
```

QY 112 AAGGGGAATTATCGACGCTGGAAGGAAGAAAGTGGTTATTCCTCCAAAGGAATAC 171
 |||||
 Db 96 AAGCGGACTAAATGATCGAGTAAGAGGGTATTTATCTTGGCGCTAACCGGTC 155
 |||||
 QY 172 GCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGTTTATGGGGATGAAG 231
 |||||
 Db 156 GCCCTAGCCCATACGATAGATGATTTTAAATGCCCTAAAGACACTTTAGCGATTTGTTTA 215
 |||||
 QY 232 ATAAGAGCAGTTTACGAAGAAGGGTTATTTGCTCTTTTGGCTTGGTTGACCCAGGA 291
 |||||
 Db 216 GGCATAAAGCATTACCTAGGTGTGGGATTTATGTAATGTTACGCTTTTGGAGCCGAA 275
 |||||
 QY 292 TGGGATGGAACCTTACACTATGCTCTACAATGCTCAATGCTCAATGAACCTGTCGAATTAGA 351
 |||||
 Db 276 TTTGAAGGCTATTATACGATTTGAATTTCTACACCACTAATTTACCGCTAAGTCTAT 335
 |||||

RESULT 10
 AAQ62300/c
 ID AAQ62300 standard; DNA; 3382 BP.

XX AC AAQ62300;

DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 27-MAY-1994 (first entry)

DE pDam-L3 plasmid fragment encoding thermostable ligase.

XX Ligase, thermostable; thermostability; Desulfohalobus ambivalens;
 KW archaeobacteria; detection; ligation; ATP; adenosine triphosphate; ss.
 XX Acidianus ambivalens.

XX Key Location/Qualifiers
 FH 1. .273
 CDS /*tag= a
 FT /note= "Open reading frame 4."
 FT complement(250. .771)
 CDS /*tag= b
 FT /note= "Open reading frame 3."
 FT 878. .883
 FT /*tag= c
 CDS 908. .2707
 FT /*tag= d
 FT /product= "Ligase."
 FT 2713. .2722
 terminator /*tag= e
 FT /note= "Potential transcription termination signal."
 FT 2804. .2811
 terminator /*tag= f
 FT /note= "Experimental transcription termination signal."
 FT 2961. .3382
 CDS /*tag= g
 FT /note= "Open reading frame 2."
 FT 2961. .3382

XX DE4217134-A1.

XX 25-NOV-1993.

XX 23-MAY-1992; 92DE-04217134.

XX 23-MAY-1992; 92DE-04217134.

XX (BOEF) BOEHRINGER MANNHEIM GMBH.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Rueger R, Kessler C, Kaletta C, Jarsch M, Kletzin A;

DR WPI; 1993-378402/48.
 DR P-PSDB; AAR44614, AAR51076, AAR51077, AAR51078.
 XX Thermotable ligase from archaeobacteria - and DNA coding for it, useful
 PT for nucleic acid detection.
 XX Claim 11; Fig 1; 26pp; German.
 XX The DNA ligase is ATP dependent and is useful for detecting nucleic acids
 CC by hybridising two oligonucleotides with adjacent sequences of the target
 CC nucleic acid, ligating the two oligonucleotides (using the ligase) and
 CC detecting the ligation product. (Updated on 25-MAR-2003 to correct PN
 CC field.) (Updated on 27-AUG-2003 to correct OS field.)
 XX Sequence 3382 BP; 1188 A; 512 C; 683 G; 999 T; 0 U; 0 Other;
 SQ Query Match 9.5%; Score 44.8; DB 2; Length 3382;
 Best Local Similarity 47.9%; Pred. No. 0.002;
 Matches 161; Conservative 0; Mismatches 172; Indels 3; Gaps 1;
 QY 120 ATTAATCGACGTGGAAAGGAAGAAAGTGGTTATTCCTCCAAAGGAATACGCTTAAT 179
 |||||
 Db 598 ATTTATGAGATAGAGAAGGAGATGAATTCATCATTTACCTAACGAACATGTACTGTT 539
 |||||
 QY 180 CCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGTTATGGGGGATATGAAGATAAGGAG 239
 |||||
 Db 538 AGTTACTGAGGATGTTAACTGCCGAACGACGTAATGGCTTTTCGTCACACTTAAGGTC 479
 |||||
 QY 240 CAGTTTAGCAAGAGAGGGTTATTTGCTCTTTTCTGGTGGTTGACCCAGGATGGATGG 299
 |||||
 Db 478 TTTCTTTGCCAGGCTGGGCTCTTTTGTACCCAGCTATCGTAGATCGAGGCTTTGAAGG 419
 |||||
 QY 300 AACTTAACACTATGCTCTACAATGCTCAATGCTCAATGAACCTGTCGAATTAAGATATGGAGA 359
 |||||
 Db 418 ACAACTCAATAGAAAGTACTAGGTTCTGCTTTT---CCCGTGAAGATAAAGAGGGGAC 362
 |||||
 QY 360 GAGATTTGTCGAGATCGCATTTATAAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGG 419
 |||||
 Db 361 TCGCTTTCTTCACTCATTTTCGCCAGAACATTACACACAGTAGAGAAATCCATTATCATGG 302
 |||||
 QY 420 AACTATATCAGGGAGCACAAAGGTTAGCGTTTTCCTTTTCAA 455
 |||||
 Db 301 CAATATCAAGGACAAACAAAGGCGTAACCTTTTACCANA 266
 |||||
 RESULT 11
 AAV90555
 ID AAV90555 standard; DNA; 2006 BP.
 XX AAV90555;
 XX 20-MAR-2003 (revised)
 DT 18-FEB-1999 (first entry)
 XX Nucleotide sequence from clone B17.
 DE Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
 XX Helicobacter pylori.
 XX WO9849314-A2.
 XX 05-NOV-1998.
 XX 25-APR-1998; 98WO-US008487.
 XX 25-APR-1997; 97US-0045107P.
 PR 14-OCT-1997; 97US-0061958P.
 XX (GENE-) GENELABS TECHNOLOGIES INC.
 XX Chow TP, Fry KE, Lim MY, Mcatee CP;

XX WPI; 1999-009433/01.
 XX New Helicobacter pylori antigens and related nucleic acid sequences -
 PT useful in serological diagnosis and protective vaccines, providing long-
 PT lasting immune response.
 XX Claim 27; Page 104; 402pp; English.
 XX The present sequence encodes a Helicobacter pylori antigenic protein that
 CC is characterised by immunoreactivity with H. pylori-positive antisera.
 CC The proteins are highly immunogenic and induce a long-lasting immune
 CC response that persists even after antimicrobial treatment. In antibody-
 CC detection assays, on sera, plasma, urine, saliva etc., they are highly
 CC sensitive and specific. The specification also describes 69 previously
 CC unrecognised immunogenic cluster families. H. pylori antigens are used to
 CC detect H. pylori-specific antibodies, for diagnosing infection or to
 CC confirm eradication of infection, and in vaccines to protect against H.
 CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
 CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
 XX Sequence 2006 BP; 690 A; 377 C; 374 G; 565 T; 0 U; 0 Other;
 SQ

Query Match 9.4%; Score 44.2; DB 2; Length 2006;
 Best Local Similarity 47.8%; Pred. No. 0.0025;
 Matches 130; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 112 AAGGGGAAATTAATCGACGTGGAAAGGAAAGGAAAGTCTTATCTCCCAAGGGAATAC 171
 DB 343 AAGCGACTAAATTAATCGACGTGGAAAGGAAAGGAAAGTCTTATCTGCGGTACGCGTTC 402

QY 172 GCCTTAATCTAACCTCGAGAGGATAAAGTTCCCGACGATGTTATGGGGGATATGAAG 231
 DB 403 GCCTTAGCTCATACGATAGATATTTTAAATGCCTAAAGACACCTTAGCGATTGTTTA 462

QY 232 ATAGAGAGAGTTTAGCAAGAGAGGGTTATTTGGTCTTCTTGGTGGTGGTACCCAGGA 291
 DB 463 GGCAAAAGCACTTACGCCAGGTGGGATCAATTTGTAATGTTACGCTTTTGGAGCCAGAA 522

QY 292 TGGGATGGAACTTAACACTAATGCTCTACAATGCCCTCAATGAACCTGTCGAATTAAGA 351
 DB 523 TTGGAAGGGTATATACGATGTAATTTCTAACACACCAACCTTACCGGCTAAAGTCTAT 582

QY 352 TATGGAGAGAGATTGTGCAGATCGCATTTATA 384
 DB 583 GCCAATGAGGGGATCGCGCAAGTGGTGTGTTTAA 615

RESULT 12
 AAV90871
 ID AAV90871 standard; DNA; 2588 BP.
 XX AAV90871;
 AC
 XX 20-MAR-2003 (revised)
 DT 18-FEB-1999 (first entry)
 XX
 XX Nucleotide sequence of cluster 28.
 XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
 XX Helicobacter pylori.
 OS
 XX WO9849314-A2.
 PN
 XX 05-NOV-1998.
 PD
 XX 25-APR-1998; 98WO-US008487.
 PF
 XX 25-APR-1997; 97US-0045107P.
 PR 14-OCT-1997; 97US-0061958P.
 XX

(GENE-) GENELABS TECHNOLOGIES INC.
 Chow TP, Fry KE, Lim MY, Mcatee CP;
 WPI; 1999-009433/01.
 XX New Helicobacter pylori antigens and related nucleic acid sequences -
 PT useful in serological diagnosis and protective vaccines, providing long-
 PT lasting immune response.
 XX Claim 22; Page 295; 402pp; English.
 XX The present sequence encodes a Helicobacter pylori antigenic protein that
 CC is characterised by immunoreactivity with H. pylori-positive antisera.
 CC The proteins are highly immunogenic and induce a long-lasting immune
 CC response that persists even after antimicrobial treatment. In antibody-
 CC detection assays, on sera, plasma, urine, saliva etc., they are highly
 CC sensitive and specific. The specification also describes 69 previously
 CC unrecognised immunogenic cluster families. H. pylori antigens are used to
 CC detect H. pylori-specific antibodies, for diagnosing infection or to
 CC confirm eradication of infection, and in vaccines to protect against H.
 CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
 CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
 XX Sequence 2588 BP; 832 A; 510 C; 486 G; 760 T; 0 U; 0 Other;
 SQ

Query Match 9.4%; Score 44.2; DB 2; Length 2588;
 Best Local Similarity 47.6%; Pred. No. 0.0027;
 Matches 130; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 112 AAGGGGAAATTAATCGACGTGGAAAGGAAAGGAAAGTCTTATCTCCCAAGGGAATAC 171
 DB 926 AACGCGACTAAATTTGTCGAGTAAAGAGGGTATTTATCTTGGCCGTAACGCGTTC 985

QY 172 GCCTTAATCTAACCTCGAGAGGATAAAGTTCCCGACGATGTTATGGGGGATATGAAG 231
 DB 986 GCCTTAGCTCATACGATAGATATTTTAAATGCCCTAAAGACACCTTAGCGATTGTTTA 1045

QY 232 ATAAGGAGCAGATTAGCAAGAGAGGGGTTATTTGGTCTTCTTGGTGGTGGTACCCAGGA 291
 DB 1046 GGCAAAAGCACTTACGCCAGGTGGGATCAATTTGTAATGTTACGCCCTTTTGGAGCCAGAA 1105

QY 292 TGGGATGGAACTTAACACTAATGCTCTACAATGCCCTCAATGAACCTGTCGAATTAAGA 351
 DB 1106 TTGGAAGGGTATATCAGATTTGAAATTTCTAACACCACTTACCGGCTAAAGTCTAT 1165

QY 352 TATGGAGAGAGATTGTGCAGATCGCATTTATA 384
 DB 1166 GCCAATGAGGGGATCGCGCAAGTGGTGTGTTTAA 1198

RESULT 13
 ADR89139
 ID ADR89139 standard; DNA; 2964 BP.
 XX ADR89139;
 AC
 XX 18-NOV-2004 (first entry)
 DT
 XX Nucleotide sequence of a lysine decarboxylase enzyme.
 DE
 XX lysine decarboxylase; enzyme; ldc gene; lysine; cadaverine; L-lysine;
 KW lysine production; gene; ds.
 XX Methylophilus methylotrophus.
 OS
 XX Key Location/Qualifiers
 FH 684..2930
 FT CDS /*tag= a
 FT /product= "lysine decarboxylase"
 XX FR2851575-A1.
 XX

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 06:11:50 ; Search time 2574 Seconds
(without alignments)
8866.514 Million cell updates/sec

Title: US-08-957-709A-70

Perfect score: 471
Sequence: 1 atgctacttccagactggaa.....caagagaaagaactcttag 471

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pi:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_uni:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	471	100.0	471	1	AY066005 Pyrococcus
2	471	100.0	471	1	AY443493 Pyrococcus
3	471	100.0	471	6	BD133238 Polymaras
4	471	100.0	471	6	AR535070 Sequence
5	471	100.0	13064	1	AE010292 Pyrococcus
6	374	79.4	253505	1	AP000007 Pyrococcus
7	316	67.1	740	6	BD133228 Polymaras
8	316	67.1	740	6	AR180965 Sequence
9	316	67.1	740	6	AR207520 Sequence
10	316	67.1	740	6	AR225643 Sequence
11	316	67.1	740	6	AR535060 Sequence
12	280.8	59.6	65118	6	BD430793_17
13	280.8	59.6	265118	1	CNSPAX06
14	280.8	59.6	265118	6	AX041922 Sequence
15	96.4	20.5	129	6	BD133227 Polymaras
16	96.4	20.5	129	6	AR180964 Sequence
17	96.4	20.5	129	6	AR207519 Sequence
18	96.4	20.5	129	6	AR225642 Sequence
19	96.4	20.5	129	6	AR535059 Sequence

C	20	90.8	19.3	600	14	AF022221	Archaeal
	21	82.8	17.6	32308	14	SVI414696	AD414696 Sulfolobu
	22	82	17.4	84	6	BD133226	BD133226 Polymaras
	23	82	17.4	84	6	AR180963	AR180963 Sequence
	24	82	17.4	84	6	AR207518	AR207518 Sequence
	25	82	17.4	84	6	AR225641	AR225641 Sequence
	26	82	17.4	84	6	AR535058	AR535058 Sequence
	27	75.8	16.1	35450	14	SVI344259	AJ344259 Sulfolobu
	28	57	12.1	110000	1	AE017261_01	Continuation (2 of
	29	56.6	12.0	75294	7	AJ783769	AJ783769 Sulfolobu
	30	54.2	11.5	7218	6	I66494	I66494 Sequence 14
	31	49.6	10.5	10698	1	AE009852	AE009852 Pyrobacul
	32	49	10.4	419	6	AX790579	AX790579 Sequence
	33	49	10.4	11390	1	AE000554	AE000554 Helicobac
	34	47.4	10.1	239340	1	AP000993	AP000993 Thermopla
	35	46.4	9.9	543	6	BD180240	BD180240 Highly th
	36	46.4	9.9	247910	1	AE017307	AE017307 Thermus t
	37	46.4	9.9	263050	1	AP000981	AP000981 Sulfolobu
	38	46	9.8	338100	1	TACID2	AL445064 Thermopla
	39	45.6	9.7	14286	1	AE000747	AE000747 Aquifex a
	40	45.4	9.6	11922	1	AE006657	AE006657 Sulfolobu
	41	45.4	9.6	349174	1	AB063522	AB063522 Wiggleswo
	42	44.8	9.5	3382	1	DALIG	X63438 D.ambivalen
	43	44.8	9.5	3382	6	A78674	A78674 Sequence 5
	44	44.2	9.4	2006	6	BD061680	BD061680 Antigenic
	45	44.2	9.4	2588	6	BD061996	BD061996 Antigenic

ALIGNMENTS

RESULT 1	AY066005	Pyrococcus furiosus dutpase gene, complete cds.	471 bp	DNA	linear	BCT 28-JAN-2002
LOCUS	AY066005	Pyrococcus furiosus dutpase gene, complete cds.				
DEFINITION	AY066005	Pyrococcus furiosus dutpase gene, complete cds.				
ACCESSION	AY066005.1	GI:18389121				
VERSION	AY066005.1	GI:18389121				
KEYWORDS		Pyrococcus furiosus				
SOURCE		Pyrococcus furiosus				
ORGANISM		Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus.				
REFERENCE		1 (bases 1 to 471)				
AUTHORS		Hogrefe,H.H., Hansen,C.J., Scott,B.R. and Nielson,K.B.				
TITLE		Archaeal dutpase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation				
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 99 (2), 596-601 (2002)				
MEDLINE		21664347				
PUBMED		11782527				
REFERENCE		2 (bases 1 to 471)				
AUTHORS		Hansen,C.J. and Hogrefe,H.H.				
TITLE		Direct Submission				
JOURNAL		Submitted (10-DEC-2001) R&D, Stratagene, 11011 N. Torrey Pines Rd., La Jolla, CA 92037, USA				
FEATURES		Location/Qualifiers				
source		1..471				
		/organism="Pyrococcus furiosus"				
		/mol_type="genomic DNA"				
		/db_xref="taxon:2261"				
CDS		1..471				
		/codon_start=1				
		/transl_table=11				
		/product="dUTPase"				
		/protein_id="AAL47572.1"				
		/db_xref="GI:18389122"				
		/translation="MLLPDWKIRKILIEPFSEESLPAGYDLRVCREAPVKGKLLDV EKEGVIPPRPYALILRLPDVMDGMRKRSLSAREGVIGSFAWDPQDGNL TMLYNASNEPVRLYGERFVQIAFIRLEGPARNPYRGNVQGSRLAFSKRKL"				
ORIGIN						
Query Match		100.0%;	Score	471;	DB 1;	Length 471;
Best Local Similarity		100.0%;	Pred. No.	4.6e-126;		
Matches		471;	Conservative	0;	Mismatches	0;
					Indels	0;
					Gaps	0;

source	1. .471				
	/organism="unidentified"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32644"				
ORIGIN					
Query Match	100.0%;	Score 471;	DB 6;	Length 471;	
Best Local Similarity	100.0%;	Pred. No. 4.6e-126;			
Matches 471;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTCTGAAGAA	60		
Db	1	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTCTGAAGAA	60		
QY	61	TCGCTCCAAACGACGAGTTATGACCTCAGAGTGGCGAGAGGCTTTTGTAAAGGGGAAA	120		
Db	61	TCGCTCCAAACGACGAGTTATGACCTCAGAGTGGCGAGAGGCTTTTGTAAAGGGGAAA	120		
QY	121	TTAATCGAGTGGAAAAGGAAAGTGGTATTCCTCAAGGGAATACGCTTTAATC	180		
Db	121	TTAATCGAGTGGAAAAGGAAAGTGGTATTCCTCAAGGGAATACGCTTTAATC	180		
QY	181	CTAACCCTCGAGAGGATAAAGTTGCCGAGCATGTTATGGGGATATGAAGATAAGGAGC	240		
Db	181	CTAACCCTCGAGAGGATAAAGTTGCCGAGCATGTTATGGGGATATGAAGATAAGGAGC	240		
QY	241	AGTTTACGAGGAGGAGGTTATTTGCTTTTGGTTGCTTCCAGGAAAGAACTCTAG	300		
Db	241	AGTTTACGAGGAGGAGGTTATTTGCTTTTGGTTGCTTCCAGGAAAGAACTCTAG	300		
QY	301	AACTTAACTAATGCTCTCAATGCTCAATGAACTGTTCGAATTAAGATATGGAGAG	360		
Db	301	AACTTAACTAATGCTCTCAATGCTCAATGAACTGTTCGAATTAAGATATGGAGAG	360		
QY	361	AGATTTGTCAGATCCATTTATAGGCTAGAGGTCGGCAAGAACCCCTTACAGAGGA	420		
Db	361	AGATTTGTCAGATCCATTTATAGGCTAGAGGTCGGCAAGAACCCCTTACAGAGGA	420		
QY	421	AACTATCAGGGAGCACAGGTTAGCGTTTCAAGAGAAAGAACTCTAG	471		
Db	421	AACTATCAGGGAGCACAGGTTAGCGTTTCAAGAGAAAGAACTCTAG	471		
RESULT 4					
LOCUS	AR535070				
DEFINITION	Sequence 70 from patent US 6734293.				
ACCESSION	AR535070				
VERSION	AR535070.1 GI:53925890				
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 471)				
AUTHORS	Hogrefe,H. and Hansen,C.J.				
TITLE	Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF protein, and methods for purifying and identifying				
JOURNAL	Patent: US 6734293-A 70 11-MAY-2004;				
FEATURES	Location/Qualifiers				
source	1. .471				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
Query Match	100.0%;	Score 471;	DB 6;	Length 471;	
Best Local Similarity	100.0%;	Pred. No. 4.6e-126;			
Matches 471;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTCTGAAGAA	60		
Db	1	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTCTGAAGAA	60		
QY	61	TCGCTCCAAACGACGAGGTTATGACCTCAGAGTGGCGAGAGGCTTTTGTAAAGGGGAAA	120		

Db	61	TCGCTCCAAACGACGAGGTTATGACCTCAGAGTGGCGAGAGGCTTTTGTAAAGGGGAAA	120	
QY	121	TTAATCGAGTGGAAAAGGAAAGTGGTATTCCTCAAGGGAATACGCTTTAATC	180	
Db	121	TTAATCGAGTGGAAAAGGAAAGTGGTATTCCTCAAGGGAATACGCTTTAATC	180	
QY	181	CTAACCCTCGAGAGGATAAAGTTGCCGAGCATGTTATGGGGATATGAAGATAAGGAGC	240	
Db	181	CTAACCCTCGAGAGGATAAAGTTGCCGAGCATGTTATGGGGATATGAAGATAAGGAGC	240	
QY	241	AGTTTACGAGGAGGAGGTTATTTGCTTTTGGTTGCTTCCAGGAAAGAACTCTAG	300	
Db	241	AGTTTACGAGGAGGAGGTTATTTGCTTTTGGTTGCTTCCAGGAAAGAACTCTAG	300	
QY	301	AACTTAACTAATGCTCTCAATGCTCAATGAACTGTTCGAATTAAGATATGGAGAG	360	
Db	301	AACTTAACTAATGCTCTCAATGCTCAATGAACTGTTCGAATTAAGATATGGAGAG	360	
QY	361	AGATTTGTCAGATCCATTTATAGGCTAGAGGTCGGCAAGAACCCCTTACAGAGGA	420	
Db	361	AGATTTGTCAGATCCATTTATAGGCTAGAGGTCGGCAAGAACCCCTTACAGAGGA	420	
QY	421	AACTATCAGGGAGCACAGGTTAGCGTTTCAAGAGAAAGAACTCTAG	471	
Db	421	AACTATCAGGGAGCACAGGTTAGCGTTTCAAGAGAAAGAACTCTAG	471	
RESULT 5				
AE010292/c				
LOCUS	Pyrococcus furiosus DSM 3638, section 167 of 173 of the complete genome.			
DEFINITION	Pyrococcus furiosus DSM 3638			
ACCESSION	AE010292.1 GI:18894190			
VERSION	AE010292.1			
KEYWORDS	Pyrococcus furiosus DSM 3638			
SOURCE	Pyrococcus furiosus DSM 3638			
ORGANISM	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus.			
REFERENCE	1 (bases 1 to 13064)			
AUTHORS	Maeder,D.L., Weiss,R.B., Dunn,D.M., Cherry,J.L., Gonzalez,J.M., Diruggiero,J. and Robb,F.T.			
TITLE	Divergence of the hyperthermophilic archaea Pyrococcus furiosus and P. horikoshii inferred from complete genomic sequences			
JOURNAL	Genetics 152 (4), 1299-1305 (1999)			
MEDLINE	99359404			
PUBMED	10430560			
REFERENCE	2 (bases 1 to 13064)			
AUTHORS	Robb,F.T., Maeder,D.L., Brown,J.R., DiRuggiero,J., Stump,M.D., Yeh,R.K., Weiss,R.B. and Dunn,D.M.			
TITLE	Genomic sequence of hyperthermophile, Pyrococcus furiosus: implications for physiology and enzymology			
JOURNAL	Meth. Enzymol. 330, 134-157 (2001)			
MEDLINE	21079003			
PUBMED	11210495			
REFERENCE	3 (bases 1 to 13064)			
AUTHORS	Weiss,R.B., Dunn,D.M., Robb,F.T. and Brown,J.R.			
TITLE	The complete sequence of the Pyrococcus furiosus genome			
JOURNAL	Unpublished			
MEDLINE	4 (bases 1 to 13064)			
PUBMED	Weiss,R.B.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (12-FEB-2002) Human Genetics, University of Utah, 20			
TITLE	South 2030 East, Salt Lake City, UT 84112, USA			
JOURNAL	Location/Qualifiers			
FEATURES	1. .13064			
source	/organism="Pyrococcus furiosus DSM 3638"			
	/mol_type="genomic DNA"			
	/strain="DSM 3638"			
	/db_xref="taxon:186497"			
	70. .573			
gene	/gene="PF1985"			

CDS	70. .573 /gene="PF1985" /note="Function Code: 16.1 Conserved Hypothetical" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAL82109.1" /db_xref="GI:18894191" /translation="WHFIYMERHYTLKASKILGVYVKTQLQNDKQGIKIRVTPGGR RSIPSEIKRILGIEHEERKVGARVSRKTQDDLRQVQAIQYAKKGNVVELLK ISSGLNEKRKSSKWMQTRKCOSSHHTSRQTNKIWIQNTQNTLSSPMNRDNNQRNSK RAKRRTH" 739. .1611 /gene="PF1986" 739. .1611 /gene="PF1986" /note="Function Code: 16.1 Conserved Hypothetical" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAL82110.1" /db_xref="GI:18894192" /translation="NKSITNMQSLKKILSTEFTKVERKQLRDLKKWNYAAHYV DSAINQMLGLVSKYRKLRRGKVRPKLKKFVYVKSFTLFTKGSVLRITIPKYY LEVINQALPYLLPFLLKELFEGKIKLGLFPDPLVLFNFKTVFVKPRDMSVDINL TNVTVLANTVNFDTRELYHIHRYVEVKROKIOKISAWNKKLSEKLKKYSRRERNK SKDFLHLANKLIVEARERGMGIILEDLNGINKENVLRGSKGLNRKLSKNWRELQRLI EYKAKWFGVPVCLC" 1975. .2388 /gene="PF1987" 1975. .2388 /gene="PF1987" /note="Function Code: 16.1 Conserved Hypothetical" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAL82111.1" /db_xref="GI:18894193" /translation="NAIGHNIIRVEFEKNVPVPGQVEVSLNPKIEELRLGEINLPNG KYKGLEVELTYNIISPEVKGKIMRIVFYLPRNREDVDNILDNWEKPKUPGELVAE VYNFANLLIPFAMMITKMGMPSPPIPRVVVQ" 2457. .3575 /gene="PF1988" 2457. .3575 /gene="PF1988" /note="Function Code: 4.2 Cellular Processes: Cell division" /codon_start=1 /transl_table=11 /product="cell division protein ftsz homolog" /protein_id="AAL82112.1" /db_xref="GI:18894194" /translation="MLKLVENVGRVSEENKVPVQVQPSIDELKKIVEQIKARI YVVGAGACNTVRMMEVGTGAKIIAVNTDAQDLLKVKHQKILIGELTRGIGAG NDPKIGEEAKESRELRLDLEGADWVFITCGLGGTGTGAAPVIAETAKRNGALTVS VYTLPTREGIRAKNAEYGLKLVKYSDFTVIIPNDKLEVAKPLIQMAFKVADEI LVQAVGITELTKGLVNLDFNRVNMKGVMIGIGESDSKRALEARQALNS PLLDVIGSATGALIHISGADVLEAAQOIIEYVRNVDPKQAQVIWGIQLEPELEKTI RVWVITGTSYITPEETPLETPEPSISISPEL" 3632. .3817 /gene="PF1989" 3632. .3817 /gene="PF1989" /note="Function Code: 4.6 Cellular Processes: Protein and peptide secretion" /codon_start=1 /transl_table=11 /product="protein transport protein sec61 gamma subunit" /protein_id="AAL82113.1" /db_xref="GI:18894195" /translation="NAELQERIRHFWKESRRAPLVTKPNWATYKRAAKITGLGIILI GLIGMLIRIVGILILGG" 3821. .4279	gene
CDS	3821. .4279 /gene="PF1990" /note="Function Code: 12.3 Transcription: RNA processing" /codon_start=1 /transl_table=11 /product="transcription antitermination protein nusG" /protein_id="AAL82114.1" /db_xref="GI:18894196" /translation="WAKKIPAVRVTHGOETAKLIYSKVRTYNLPVAILAPSRVKG YIFVEAPNGVVDRAIRGIRHARGVLPGEVFPKIEHFELEKPAVSGLEPGDLVEIA GPFQKGAKVVKIDESKDEVVVQFDALIVPVIKGDYVRLISLQKEE" 4291. .4785 /gene="PF1991" 4291. .4785 /gene="PF1991" /note="Function Code: 13.4 Translation: Ribosomal proteins: synthesis and modification; (rpl11p)" /codon_start=1 /transl_table=11 /product="LSU ribosomal protein L11p" /protein_id="AAL82115.1" /db_xref="GI:18894197" /translation="MPKQVVELVEGGKATGPPGLGPAIGPLGLNVKQVVDKINEATK DFAGMQVPKIIIVPVTQFEIEVGPPTSQIIRKLELGKSGSGPKHNVGNLTMEQ VIKIAKMKDQMLALTAKAAKEVIGTALSMGTVVEGKDPREVQKEIDEGVYDELFEK AEKS" 4864. .5514 /gene="PF1992" 4864. .5514 /gene="PF1992" /note="Function Code: 13.4 Translation: Ribosomal proteins: synthesis and modification; (rpl11p)" /codon_start=1 /transl_table=11 /product="LSU ribosomal protein L1p" /protein_id="AAL82116.1" /db_xref="GI:18894198" /translation="MPPDROKIVKAKKAKARAKPRNFTQSVEAVNLKIDILKRPEN RFLKLVLPHGKGDVIAVADGAAVEAAKGLGDISSAELEBIASSPROAKLAK KYDFIAEAPLMPKIGRYLGRNPNPVPVPTLTDLPIVEKLAKTVPIQIKNN PVVHAPVGTETKMSDEIAENIAVLNAILGKERGESQVKSYYVKTMTGPAVKIES" 5520. .6539 /gene="PF1993" 5520. .6539 /gene="PF1993" /note="Function Code: 13.4 Translation: Ribosomal proteins: synthesis and modification; (rpl10e)" /codon_start=1 /transl_table=11 /product="LSU ribosomal protein L10E" /protein_id="AAL82117.1" /db_xref="GI:18894199" /translation="MAHVAEWKKEVEBELANLIKSPVVALVDVSSMPAYPLSQMRRLL IRENGLLRVRNTLIELAIKKVAGSELKPELEKLIYEGGAGILVTMNPFLKFKF LQONQRPAPKPGAKVPKDVIPAGTSLAPGIVGQMGAPARIERKVTIOKDT VLKAGEBITELANILNALGIQPLEVGLDLAVVEDGIIYTPDVLAIDESEYINMLQ KAYHAFNLVDNIAYPTQTIEALIQKALFNKAVAVAGYITKETSIDIIGRAIRAM LAILQQLPDELVDLDEKTKELLSAQAVSQVSEEEKKEEKEEKEEKEEKEEALAGL SALFG" 6586. .6909 /gene="PF1994" 6586. .6909 /gene="PF1994" /note="Function Code: 13.4 Translation: Ribosomal proteins: synthesis and modification; (rpl12a)" /codon_start=1 /transl_table=11 /product="LSU ribosomal protein L12a" /protein_id="AAL82118.1" /db_xref="GI:18894200" /translation="MEYVIAALLHSVGKEINEENLKAVALQAGVPEPPARIKALVAA LEGVNIDEVIEKAMPVAVAAAAAPAGGGGEEKKEEKEEKEEKEEVEEVEEALAGL	gene


```
gene          SALFG"
7014. 7556
/genes="PF1995"
CDS           7014. 7556
/genes="PF1995"
/notes="Function Code: 16.1 Conserved Hypothetical"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAU82119.1"
/db_xref="GI:18894201"

Query Match      100.0%; Score 471; DB 1; Length 13064;
Best Local Similarity 100.0%; Pred. No. 5.1e-126;
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60
DB ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 7947
QY 61 TCGCTCAACACGAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTGTAAAGGGAAA 120
DB TCGCTCAACACGAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTGTAAAGGGAAA 7887
QY 121 TTAATCGAGTGGAAAGGAAGGAAGTGGTTATTCCTCCAGGGAATACGCTTAATC 180
DB TTAATCGAGTGGAAAGGAAGGAAGTGGTTATTCCTCCAGGGAATACGCTTAATC 7827
QY 181 CTAAACCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
DB CTAAACCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 7767
QY 241 AGTTTACGAGAGGAGGTTATGTTCTTTTCTTGGTTTCCACGAGATGGGATGGA 300
DB AGTTTACGAGAGGAGGTTATGTTCTTTTCTTGGTTTCCACGAGATGGGATGGA 7707
QY 301 AACTTAACACTAATGCTTACATGCTCAATGAACCTGTCAATTAAGATATGAGAG 360
DB AACTTAACACTAATGCTTACATGCTCAATGAACCTGTCAATTAAGATATGAGAG 7647
QY 361 AGATTTGTGCGAGATCCATTTATAAGCTAGAGGTCGGCAAGAAACCTTACAGAGGA 420
DB AGATTTGTGCGAGATCCATTTATAAGCTAGAGGTCGGCAAGAAACCTTACAGAGGA 7587
QY 421 AACTATCAGGGGAGCACAAGGTTAGCGTTTTCAAAGAGAAAGAACTCTAG 471
DB 7586 AACTATCAGGGGAGCACAAGGTTAGCGTTTTCAAAGAGAAAGAACTCTAG 7536

RESULT 6
AP000007 Pyrococcus horikoshii OT3 DNA linear BCT 27-MAY-2004
LOCUS     253505 bp
DEFINITION Pyrococcus horikoshii OT3 DNA, complete genome, 1485001-1738505 nt.
POSITION, section 7/7.
ACCESSION AP000007 AB009464 AB009465 AB009521 AB009522 AB009523 AB009524
AB009525 AB009526 AB009527 AB009528 AB009529 AB009530 AB009531
BA000001
VERSION   AP000007.1 GI:3236134
KEYWORDS  Pyrococcus horikoshii OT3
SOURCE    Pyrococcus horikoshii OT3
ORGANISM  Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
REFERENCE 1
AUTHORS   Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y.,
Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y.,
Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Onofuku, Y.,
Funahashi, T., Tanaka, T., Kudo, Y., Yamazaki, J., Kushiida, N.,
Oguchi, A., Aoki, K., Nakamura, Y., Robb, T.F., Horikoshi, K.,
Masuchi, Y., Shizuya, H. and Kikuchi, H.
Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3
DNA Res. 5 (2), 55-76 (1998)

JOURNAL
```

```
98344137
9679194
2 (bases 1 to 253505)
Director-General of Biotechnology Center, Tanaka, T.,
Kawarabayasi, Y. and Kikuchi, H.
Direct Submission
Submitted (11-JUN-1998) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan
(E-mail: bio@nitech.go.jp, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
On or before Mar 17, 1999 this sequence version replaced
gi:3131896, gi:3131920, gi:3131974, gi:3131982, gi:3132032,
gi:3132063, gi:3132118, gi:3132139, gi:3132184, gi:3132233,
gi:3133142.
Kawarabayasi, Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Robb, T. F. is at the Center of Marine Biotechnology, University of
Maryland, Baltimore, MD, USA.
Horikoshi, K. is at the Japan Marine Science and Technology Center,
Yokosuka, Kanagawa 237-0061, Japan.
Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
Japan.
Shizuya, H. is at the California Institute of Technology, Pasadena,
CA, USA.
The other authors are at the National Institute of Technology and
Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG
and stop codon are defined as CDS
Homology analysis is performed by Smith-Waterman algorithm against
GenBank and GenPept release 103; EMBL release 52.0; SwissProt
release 34.0; PIR-Protein release 54.0; and OWL release 29.5.
E-mail address for comments and questions: genome@nitech.go.jp
Restriction map, ORF organization, sequence alignment and more
information are available at W.W.W. site of Biotechnology Center,
URL: http://www.bio.nitech.go.jp/.

FEATURES
Location/Qualifiers
1..253505
/organism="Pyrococcus horikoshii OT3"
/mol_type="genomic DNA"
/strain="OT3"
/db_xref="taxon:70601"
/map="1485001-1738505 nt. position"

ORIGIN
Query Match      79.4%; Score 374; DB 1; Length 253505;
Best Local Similarity 87.2%; Pred. No. 9.9e-98;
Matches 410; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60
DB 250855 ATGCTACTTCCGAGCTGGAAATAGAAAGAAATACTTATAGAACCAITTTCTGAAGAA 250914
QY 61 TCGCTCAACACGAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTGTAAAGGGAAA 120
DB 250915 TCATTTCAACACGAGGTTATGATCTTAGAGTGGGTAGAGAGCTTTTGTAGTGAAG 250974
QY 121 TTAATCGAGTGGAAAGGAAGGAAGTGGTTATTCCTCCAGGGAATACGCTTAATC 180
DB 250975 TTAATGTAGTGGAAAGGAAGGAAGTGGTTATTCCTCCAGGGAATACGCTTAAT 251034
QY 181 CTAAACCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
DB 251035 CTAAACCTCGAGAGGATAAAGTTACCTGATGATGATGAGATATGAAGATAAGGAGC 251094
QY 241 AGTTTACGAGAGGAGGTTATGTTCTTTTCTTGGTTTCCACGAGATGGGATGGA 300
DB 251095 AGCTTAGCTAGAGAGGGGATTCAGGTTCTTTTTCCTGGGTGACCCAGGATGGGATGG 251154
QY 301 AACTTAACACTAATGCTTACATGCTCAATGAACCTGTTCGAAATTAAGATATGAGAG 360
DB 251155 AACTTAACCTTAATGCTTACATGCTCAATGAACCAAGTAGAATTAATACGAGAA 251214
```

QY 361 AGATTGTGCAGATCGCATTTATTAAGGCTAGAGGTCCGCAAGAAACCCCTTACAGAGCA 420
|||
Db 251215 AGTTTGTGCAGATTGTCTTTATAGGCTAGAGGATCCACCAAGGAATCCCTATAGTGA 251274
|||
QY 421 AACTATCAGGGAGCAACAGGTAGGTTTTCAGAGAGAAAGAACTCTA 470
|||
Db 251275 AACTATCAGGAAGCAACAGATTGGTATTTTCAAGAGAAAGAACTCTA 251324
|||
RESULT 7
BD133228 740 bp DNA linear PAT 18-SEP-2002
LOCUS
DEFINITION Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
isolated PEF proteins, and methods for purifying and identifying
them.
ACCESSION BD133228
VERSION BD133228.1 GI:23228173
KEYWORDS JP 2002505572-A/24.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe,H. and Hansen,C.J.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
isolated PEF proteins, and methods for purifying and identifying
JOURNAL Patent: JP 2002505572-A 24 19-FEB-2002;
STRATAGENE
COMMENT OS Unidentified
PN JP 2002505572-A/24
PD 19-FEB-2002
PF 20-MAR-1998 JP 1998545828
PR 21-MAR-1997 US 08/822774,24-OCT-1997 US 08/957709 PI
HOLLY HOGREFE,CONNIE J HANSEN
PC Cl2P19/34,Cl2Q1/68,Cl2N9/12
CC Strandedness: Single;
CC Topology: Linear;
CC Polymerase enhancing factor (PEF) extracts, PEF protein CC
complexes.
CC isolated PEF proteins, and methods for purifying and CC
identifying them
FH Key Location/Qualifiers
FT source 1..740
FT /organism='Unidentified'.
FEATURES
source Location/Qualifiers
1..740
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGATAAAGTTGCCCGACGATGT 215
|||
Db 1 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGATAAAGTTGCCCGACGATGT 60
|||
QY 216 TATGGGGATATGAAGATGAAGAGAGATTTAGCAAGAGAGGGGTATTGGTTCTTTTGC 275
|||
Db 61 TATGGGGATATGAAGATGAAGAGAGATTTAGCAAGAGAGGGGTATTGGTTCTTTTGC 120
|||
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCTCAAATGA 335
|||
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCTCAAATGA 180
|||
QY 336 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 395
|||
Db 181 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 240
|||
QY 396 TCCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCAAGGTTAGCGTTTTCAAA 455
|||
Db 241 TCCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCAAGGTTAGCGTTTTCAAA 300
|||
Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGATAAAGTTGCCCGACGATGT 215
|||
Db 1 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGATAAAGTTGCCCGACGATGT 60
|||
QY 216 TATGGGGATATGAAGATGAAGAGAGATTTAGCAAGAGAGGGGTATTGGTTCTTTTGC 275
|||
Db 61 TATGGGGATATGAAGATGAAGAGAGATTTAGCAAGAGAGGGGTATTGGTTCTTTTGC 120
|||
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCTCAAATGA 335
|||
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCTCAAATGA 180
|||
QY 336 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 395
|||
Db 181 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 240
|||
QY 396 TCCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCAAGGTTAGCGTTTTCAAA 455
|||
Db 241 TCCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCAAGGTTAGCGTTTTCAAA 300
|||

QY 456 GAGAAAGAACTCTAG 471
|||
Db 301 GAGAAAGAACTCTAG 316
|||
RESULT 8
AR180965 740 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 42 from patent US 6333165.
ACCESSION AR180965
VERSION AR180965.1 GI:20222998
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
unclassified.
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe,H.
TITLE Methods for identifying polymerase enhancing factor (PEF)
JOURNAL Patent: US 6333165-A 42 25-DEC-2001;
FEATURES Location/Qualifiers
1..740
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGATAAAGTTGCCCGACGATGT 215
|||
Db 1 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGATAAAGTTGCCCGACGATGT 60
|||
QY 216 TATGGGGATATGAAGATGAAGAGAGATTTAGCAAGAGAGGGGTATTGGTTCTTTTGC 275
|||
Db 61 TATGGGGATATGAAGATGAAGAGAGATTTAGCAAGAGAGGGGTATTGGTTCTTTTGC 120
|||
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCTCAAATGA 335
|||
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCTCAAATGA 180
|||
QY 336 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 395
|||
Db 181 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 240
|||
QY 396 TCCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCAAGGTTAGCGTTTTCAAA 455
|||
Db 241 TCCGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCAAGGTTAGCGTTTTCAAA 300
|||
QY 456 GAGAAAGAACTCTAG 471
|||
Db 301 GAGAAAGAACTCTAG 316
|||
RESULT 9
AR207520 740 bp DNA linear PAT 20-JUN-2002
LOCUS
DEFINITION Sequence 42 from patent US 6379553.
ACCESSION AR207520
VERSION AR207520.1 GI:21507294
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
unclassified.
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe,H.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
isolated PEF proteins, and methods for purifying and identifying
same
JOURNAL Patent: US 6379553-A 42 30-APR-2002;
FEATURES Location/Qualifiers
1..740
source

/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 TCCTCCAAAGGGAATACGCCCTTAATCCTTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 215
DB 1 TCCTCCAAAGGGAATACGCCCTTAATCCTTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 60

QY 216 TATGGGGGATATGAAGTATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 275
DB 61 TATGGGGGATATGAAGTATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 120

QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAAAATGA 335
DB 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAAAATGA 180

QY 336 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 395
DB 181 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 240

QY 396 TCCGGCAAGAAACCCCTTACAGAGAACTATCAGGGAGGACAAAGTTAGCGTTTTCAAA 455
DB 241 TCCGGCAAGAAACCCCTTACAGAGAACTATCAGGGAGGACAAAGTTAGCGTTTTCAAA 300

QY 456 GAGAAAGAACTCTAG 471
DB 301 GAGAAAGAACTCTAG 316

RESULT 10
AR225643
LOCUS AR225643 740 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 42 from patent US 644428.
ACCESSION AR225643
VERSION AR225643.1 GI:27263675
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe,H. and Hansen,C.J.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying same
JOURNAL Patent: US 644428-A 42 03-SEP-2002;
FEATURES
source Location/Qualifiers
1..740
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 TCCTCCAAAGGGAATACGCCCTTAATCCTTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 215
DB 1 TCCTCCAAAGGGAATACGCCCTTAATCCTTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 60

QY 216 TATGGGGGATATGAAGTATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 275
DB 61 TATGGGGGATATGAAGTATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 120

QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAAAATGA 335
DB 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAAAATGA 180

QY 336 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 395
DB 181 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 240

QY 396 TCCGGCAAGAAACCCCTTACAGAGAACTATCAGGGAGGACAAAGTTAGCGTTTTCAAA 455
DB 241 TCCGGCAAGAAACCCCTTACAGAGAACTATCAGGGAGGACAAAGTTAGCGTTTTCAAA 300

QY 456 GAGAAAGAACTCTAG 471
DB 301 GAGAAAGAACTCTAG 316

RESULT 11
AR535060
LOCUS AR535060 740 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 42 from patent US 6734293.
ACCESSION AR535060
VERSION AR535060.1 GI:53925880
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe,H. and Hansen,C.J.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF protein, and methods for purifying and identifying
JOURNAL Patent: US 6734293-A 42 11-MAY-2004;
FEATURES
source Location/Qualifiers
1..740
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 TCCTCCAAAGGGAATACGCCCTTAATCCTTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 215
DB 1 TCCTCCAAAGGGAATACGCCCTTAATCCTTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 60

QY 216 TATGGGGGATATGAAGTATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 275
DB 61 TATGGGGGATATGAAGTATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 120

QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAAAATGA 335
DB 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAAAATGA 180

QY 336 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 395
DB 181 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 240

QY 396 TCCGGCAAGAAACCCCTTACAGAGAACTATCAGGGAGGACAAAGTTAGCGTTTTCAAA 455
DB 241 TCCGGCAAGAAACCCCTTACAGAGAACTATCAGGGAGGACAAAGTTAGCGTTTTCAAA 300

QY 456 GAGAAAGAACTCTAG 471
DB 301 GAGAAAGAACTCTAG 316

Sequence split into 18 fragments LOCUS BD430793 Accession BD430793

Fragment Name Begin End
BD430793_00 1 110000
BD430793_01 100001 210000
BD430793_02 200001 310000
BD430793_03 300001 410000
BD430793_04 400001 510000
BD430793_05 500001 610000
BD430793_06 600001 710000

RESULT 12
BD430793.17
WPCOMMENT

BD430793_07	700001	810000	BD430793 GENOME SEQUENCE AND POLY			
BD430793_08	800001	910000				
BD430793_09	900001	1010000				
BD430793_10	1000001	1110000				
BD430793_11	1100001	1210000				
BD430793_12	1200001	1310000				
BD430793_13	1300001	1410000				
BD430793_14	1400001	1510000				
BD430793_15	1500001	1610000				
BD430793_16	1600001	1710000				
BD430793_17	1700001	1765118				
Continuation (18 of 18) of BD430793 from base 1700001						
Query Match 59.6%; Score 280.8; DB 6; Length 65118;						
Best Local Similarity 75.0%; Pred. No. 1.3e-70;						
Matches 351; Conservative 0; Mismatches 117; Indels 0; Gaps 0;						
Qy	1	ATGCTACTTCCAGACTGGAATAACAGAAAGAAATACCTTATAGAGCATTCTTCTGAGGAA	60			
Db	62045	ATGCTCTTCCAGACTGGAATAAAGGAAAGAGATTTTAAATAAGCCATTCTCAGAGGAA	62104			
Qy	61	TGCTCCAACCCAGCGGTATGACCTCAGATGGCGAGAGCGCTTTGTTTAAAGGGAAA	120			
Db	62105	TGCTTTCAACCCAGCTGGCTACGACTTAAGGTAGGCAAGGAAGCTTACATTCAGGGAAAG	62164			
Qy	121	TTAATCGACGTGGAAAAGGAAGGAAGTTCGTTATTCCTCCAAAGGAATACGCCCTTAATC	180			
Db	62165	TTCATAGATGTTGAAAGAGGGGCANAGTCATATACCTCCAAAGAAATATGCCCTAATA	62224			
Qy	181	CTAACCCCTCGAGAGGATAAAGTTCGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC	240			
Db	62225	CTGACCCCTAGAGAGGATAAAGCTTCCAGATGATATTATGGGAGACATGAAGATAAGGAGC	62284			
Qy	241	AGTTTAGCAAGAGAGGGGTATTGCTTCTTTGCTTGGGTGACCCAGGATGGGATGGA	300			
Db	62285	AGTCTAGCTAGGGAAGGTGTTCTAGGCTCTTTTCATGGGTAGACCCGGGATGGGACGGT	62344			
Qy	301	AACTTAACACTAATGCTCTCAATAGCCTCAATGACCTGTCGAATTAAGATATGAGAG	360			
Db	62345	AATCTAACCTTGATGCTTTACATGTCATCGGAAAAGGAGGTAAATTTTAAAGGTACAAAGAG	62404			
Qy	361	AGATTGTGCGATCGCATTTTATAAGGCTAGAGGGTCCGGCAAGAAACCTTCACAGAGGA	420			
Db	62405	AGGTTTGTCCAGATAGCTTCTTCAAGGCTTGAGGCTCCCGCCAAAGATCCATACAGGGGC	62464			
Qy	421	AACATACAGGGGACACAGGTTAGCGTTTTTCAAAGAGAAAGAACTC	468			
Db	62465	AACATCAAGGAAGCAGAGAATAGTCTCTCAAAGAGATCTTAACTC	62512			
RESULT 13						
CNSPAX06 265118 bp DNA circular BCT 29-MAY-2003						
LOCUS Pyrococcus abyssi complete genome; segment 6/6.						
DEFINITION AJ248288 AL096836						
ACCESSION AJ248288.1 GI:5458960						
VERSION complete genome.						
KEYWORDS Pyrococcus abyssi						
SOURCE Pyrococcus abyssi						
ORGANISM Archaea; Euryarchaeota; Thermococci; Thermococcales;						
Thermococcaceae; Pyrococcus.						
REFERENCE 1 (bases 1 to 265118)						
AUTHORS Gaspin,C., Cavaille,J., Erauso,G. and Bachelierie,J.P.						
TITLE Archaeal homologs of eukaryotic methylation guide small nucleolar						
RNAs: lessons from the Pyrococcus genomes						
J. Mol. Biol. 297 (4), 895-906 (2000)						
JOURNAL 20202371						
MEDLINE 10736225						
PUBMED						
REFERENCE 2						
AUTHORS Lecompte,O., Ripp,R., Puzos-Barbe,V., Duprat,S., Heilig,R.,						
Dietrich,J., Thierry,J.C. and Poch,O.						
TITLE Genome evolution at the genus level: comparison of three complete						
genomes of hyperthermophilic archaea						

JOURNAL	Genome Res. 11 (6), 981-993 (2001)
MEDLINE	21275479
PUBMED	11381026
REFERENCE	3
AUTHORS	Cohen,G., Barbe,V., Flament,D., Galperin,M., Heilig,R., Ripp,R., Lecompte,O., Prieur,D., Poch,O., Quellerou,J., Thierry,J.C., Van der Oost,J., Weissenbach,J., Zivanovic,Y. and Forterre,P.
An integrated analysis of the genome of the hyperthermophilic archaeon Pyrococcus abyssi	
TITLE	Mol. Microbiol. 47 (6), 1495-1512 (2003)
JOURNAL	22511545
MEDLINE	22622808
PUBMED	4 (bases 1 to 265118)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (06-JUL-1999) Genoscope - Centre National de Sequençage :
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)	
Join(AJ248283.1:1..307150,AJ248284.1:51..293250,	
AJ248285.1:51..307120,	
AJ248286.1:51..288440,AJ248287.1:51..304290,AJ248288.1:51..265118).	
Location/Qualifiers	
1..265118	
/organism="Pyrococcus abyssi"	
/mol_type="genomic DNA"	
/strains="Orsay"	
/db_xref="taxon:29292"	
164..1615	
/gene="bgal"	
164..1615	
/gene="bgal"	
/note="PAB2376"	
/codon_start=1	
/transl_table=11	
/product="bgal-2 beta-galactosidase (EC 3.2.1.23)	
(lactase)"	
/protein_id="CAB50447.1"	
/db_xref="GI:5458961"	
/db_xref="GOA:Q9UYG5"	
/db_xref="UniProt/TREMBL:Q9UYG5"	
/translation="NEFWGVVQSAPQEPMDPPYRNLDTRSDMHWVDPYNIKNEL	
VSGDLPESGINNYELSIDRLAKSLGNAYQLTIENSRIFPCPYSEVEVERDQYG	
FIKSVIKERLEKDELANKREVHYENLVNKLKLNFTFTVLNQHNPNIWLHDPPI	
EVRNIEKARHGVDERRAIVFESKFAAYVAMKFDKYDFWATFDEPMVTLGYLAP	
VWGPPEGLNPKAAKSVIMHQIIAHARAYDSIKKFSKDPGVILNIIIPAYLPNRDSK	
HVKAENYDLPNRLFLKLVNKGKLDIDINGEYKVPKPLKRDNDWIGNNYTREVVKYV	
EPKYKELPLVTFPVEGEGYSANPNSISPDNNPTSDRGWEVYPKGLYDSTAAYEVE	
RPFITENGADSKDILRPRIIVDHVKEVKRLRENGINVLGYFHNALTDYENWAMGPKI	
RFLYEVDPITKRIIPRRKSVETIKKVKVBGIE"	
1834..2283	
/note="PAB1019"	
/codon_start=1	
/transl_table=11	
/product="Predicted nucleic acid-binding protein, contains	
PIN domain"	
/protein_id="CAB50448.1"	
/db_xref="GI:5458962"	
/db_xref="UniProt/TREMBL:Q9UYG4"	
/translation="NPLPDPITFDSVTLLKWHSEKNKRLLEITLAKFTVYLSMLTVQR	
VLTARAYLKNRVLEFVKDIYNIVPVSEELIRASKLEANLIRKGVFLDLEIITA	
VTAIETGLLITDEPKYEPTRRFGLDTMPLDKFLREVELMVEKEII"	
complement(2276..3079)	
/note="PAB1341"	
/codon_start=1	
/transl_table=11	
/product="Hypothetical protein"	
/protein_id="CAB50449.1"	
/db_xref="GI:5458963"	
/db_xref="UniProt/TREMBL:Q9UYG3"	
/translation="MFENLESPEEAPKLLLEVLDFDLNPFDPKVEKVLCEIHPDPDC	
VIGMGTTIKLTERGIEVYACMTDGYMGTLDSSLTGHELATIRREBERESSKLLGVK	
KIWLNYRDTLPIRESREVRKDLVRIIRKPKPGVFLPDPWLPIEAPHDPHNTGFLAD	
AVAFGLPFPNSVDVEIGLGHQVFSFIADYITNKPNYFVDITDVMELKUKAIATHKSO	

```
CDS
3192..4340
/notes="PAB1020"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50450.1"
/db_xref="GI:5458964"
/db_xref="GOA:Q9UYG2"
/db_xref="UniProt/TREMBL:Q9UYG2"
/translations="MKEVSSVYKEILVKLGLEDRIETLEMKGGIIEDEFDGIRVVR
FKDSACKLRGTVIDBEVVIQGFPHIKRIINLRSGIRRIFKRGEPFVVEKVDGYNVR
VVMYKMGIGITGGICPTTIERIPDFVQPEFFKDNPNLILVGENAGPESPVLVGGP
PYKDIQFLFDVQIKTGRSLPVEERLUKIAEYGINHVEFGTKDDVDLYSLAI
ERLKGREGIKMSDKIKVIKVPYANINDIKIGARVYELPGYFTSRISRLAF
YLAEKIKGEFERVAKEIGSALLQPFVESIFDVEQEDIHFLKVRVKRIETAYKVV
THEKLGKIEIVDIEIKDGRWITFKRLYPDATNEIRELIGKAFVD"
4351..5658
/notes="PAB1021"
/codon_start=1
/transl_table=11
/product="Putative hydroxypyruvate reductase"
/protein_id="CAB50451.1"
/db_xref="GI:5458965"
/db_xref="UniProt/TREMBL:Q9UYG1"
/translations="MEDIALRVKAIARAADPVYKAVINSVRVEGNKLLVKDKFEFKIG
KYVLAFGAACDMARAEIILVEDGVAIVKYGKELKKIRVI BAGHPIDPENSIR
GATGKLEKKEEDIVFLISGGGSAFLPEPDGIGISLEDLKTNELLLKSGAKIHE
INTVRKHSKVGKGLAKIKGTGIVLITSDVVGNLEALASGPTVKDPTTFODAKRI
LELXGIDRVPSVRSRHSIEKINGEAEETLKEELNHNPLIASNISISQAVVEATK
LGFKATWITITLGEARADKGFISVQEIATKGNFPFPPVLVFGETITVTEGGGG
KGGPNELASATRKIKGLNALVAFDTDGTGPTDAAGGIVDGETYEKLRRKGIDIE
KVLKENSYEALKKVGSLLFTGPTGTNVNMIITAIITSQGR"
complement(5633..5655)
/notes="PAB1338"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50452.1"
/db_xref="GI:5458966"
/db_xref="UniProt/TREMBL:Q9UYG0"
/translations="MPIYEARVKVIGISSNVTSKVSRDALAQLESDRIVKVLNMG
VTEGDFPFSALDITGIRIDPELKNPKDHFGLKFFVPGRGELLVIFTIEENLLIYD
EKRFABFVKFEVLINGKTVMQLARIIVGGAINMESKWEGBLWIKVKSARTORT
RSVVITQDRPVYSIFDSEDIIEEVENMGKVRWKIRHPHINOSVTSYLIYIPDK
VOLVRYLLKLYTPSAMEFIKIADDPFTLKSEFOLMEKELSELSDBMEKQILVA
LYSGINPLEHLQPLGITEKEIEBIIYDMIDKGLLKIVIRKIVDLTNEGRKIVNLLK
YGLVSM"
complement(6670..7425)
/notes="PAB1337"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50453.1"
/db_xref="GI:5458967"
/db_xref="UniProt/TREMBL:Q9UYF9"
/translations="MDVGTLLIIVAVLIGISVSAWISVSIQQMKNMQALMISRLS
EVETKVKRKTGRSEAKPEPEAKQEVDRDVEEDIVRKIKOIDEKVHNVANEAKK
KKERLMLLDVARGYALGVSEDEYNALFMRVLSELDEPKRLWLAAYPPEDKRLNQ
MLTHVARTKLPVIRBKGKSGSVKLSFDEALIRITSSIKSAISILDDLIKKRGDNP
AVTFLBKLSQBELEKCKKVEDLEKRLQBLEAL"
complement(7468..7967)
/notes="PAB1336"
/codon_start=1
/transl_table=11
/product="Methyl accepting chemotaxis protein"
/protein_id="CAB50454.1"
/db_xref="GI:5458968"
/db_xref="GOA:Q9UYF8"
/db_xref="UniProt/TREMBL:Q9UYF8"
/translations="MQFRKKIILMAISAVITWAIKGGVTIISFNKVKYGVIKGEIQAP
AIEQGGQLAABAAEAKMPDFDFARVADYGTWATEVVEKAYAGLSDNETAMREFLL
SEFPKKIKELNKDVSYYVFGDEEGHMYMPPDEPGYDPRVRPWYVQAKEKGAIIIE
```

```
CDS
complement(9704..10195)
/notes="PAB1335"
/codon_start=1
/transl_table=11
/product="cheD chemotaxis protein cheD"
/protein_id="CAB50455.1"
/db_xref="GI:5458969"
/db_xref="UniProt/TREMBL:Q9UYF7"
/translations="WTRREIKVGIGDYAVKGEGIIISTYGLGSCVGITLYDRVTKVGL
LHALLPEAARYGHRGNPAKYVDVTGLQLLKVLLKGLGASKPRLEAKLFGGAOMFONIKS
DELKIGERNVQAKRELKGLIRLVAEDTGGRGRTIYLDLSTGKVRMRKVIQGVIE
KVY"
complement(10179..10817)
/notes="PAB1334"
/codon_start=1
/transl_table=11
/product="cheC-1 chemotaxis protein cheC"
/protein_id="CAB50456.1"
/db_xref="GI:5458970"
/db_xref="UniProt/TREMBL:Q9UYF6"
/translations="MKSELYKIDIFKESNIAMSHALTSLSQMIIGPIEMEADPVEIV
SRABFLUKLAERGVSXGFTVMDFVTEGLSLTILQPKHSALNISAVLMGEFGSMBE
LDMEGKSATMEVGNILISAYTDILSNLIGEPVSLSPKPAESLYDIEKELGRDLRNV
Query Match 59.6%; Score 280.8; DB 1; Length 265118;
Best Local Similarity 75.0%; Pred. No. 1.4e-70; Indels 0; Gaps 0;
Matches 351; Conservative 0; Mismatches 117;
QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAAGAAATACTATTATAGAGCCATTTCTGAAGAA 60
DB 262045 ATGCTCTTCCAGACTGGAAATCAGAAAAGAGATTTTATAAGCCATCTCAGAGAA 262104
QY 61 TCGCTCAACACGACGATTATGACCTCAGATGGGAGAGAGCGTTTGTATTAGGGGAAA 120
DB 262105 TCGCTTCAACACGCTGGCTACGACTTAAGGAGGCAAGAGCTTACATTCAGGGAAG 262164
QY 121 TTAATCGAGTGGAAAGGAGGAAAGTCGTTTATCTCCCAAGGGAATACGCTTAATC 180
DB 262165 TTCTATAGATGTTGAAAGGAGGCAAGATCATTAATACCTCCAAAGAAATATGCCCTAATA 262224
QY 181 CTAAACCTCGAGAGGATAAGTTGCCCGACGATGTTATGGGGATATGAAGATAAGGAGC 240
DB 262225 CTGACCTTAGAGAGGATTAAGCTTCCAGATGATATTATGGGACATGAAGATAAGGAGC 262284
QY 241 AGTTTAGAAGAGAGAGGGGTATTGTTCTTTTGTGGTTCACCCAGAGATGGGATGGA 300
DB 262285 AGCTAGCTAGGAGAGGTGTTCTAGGCTCCTTTGTCATGGGTAGACCCGGGATGGACGGT 262344
QY 301 AACTTAACACTAATGCTCTACATATGCTCAATGAACCTGTCGAATTAAGATATGGAG 360
DB 262345 AATCTAACCTTGATGCTCTTACATATGCTCGAAAGAGAGGATAATTTTAAAGGTACAAG 262404
QY 361 AGATTTGTGCAGATCCGATTTATTAAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAG 420
DB 262405 AGGTTTGTCCAGATAGCTTCTTAAGGCTTGAGGCTCCCGCAAGAAATCCATACAGGGG 262464
QY 421 AACTATCAGGGGAGCACAAGGTTTACGCTTTTCAAGAGCAAGAAACCTC 468
DB 262465 AACTATCAAGGAAGCAGGAGNATAGTCTCTCTCAAGAGAGATCTTAACTC 262512
RESULT 14
AX041922
```

LOCUS	AX041922	265118 bp	DNA	linear	PAT 23-NOV-2000
DEFINITION	Sequence 817 from Patent WO0065062.				
ACCESSION	AX041922				
VERSION	AX041922.1	GI:11340687			
KEYWORDS	Pyrococcus abyssi				
SOURCE	Pyrococcus abyssi				
ORGANISM	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus.				
REFERENCE	1				
AUTHORS	Forterre,P., Thierry,J.C., Prieur,D., Dietrich,J., Lecompte,O., Querellou,J., Weissenbach,J., Saurin,W., Heilig,R., Flament,D., Raffin,J.P., Henneke,G., Gueguen,Y. and Rolland,J.L.				
TITLE	Genome sequence and polypeptides of pyrococcus _i(abyssi), fragment and uses thereof				
JOURNAL	Patent: WO 0065062-A 817 02-NOV-2000; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IFREMER INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)				
FEATURES	Location/Qualifiers				
source	1..265118 /organism="Pyrococcus abyssi" /mol_type="unassigned DNA" /db_xref="taxon:29292" /notes="Original length of seq 1: 1765118 (for 812 sequences) New sequences : seq 001, from 0.000.001 to 0.349.980, length: 349.980 seq 813, from 0.300.001 to 0.649.980, length: 349.980 seq 814, from 0.600.001 to 0.949.980, length: 349.980 seq 815, from 0.900.001 to 1.249.980, length: 349.980 seq 816, from 1.200.001 to 1.549.980, length: 349.980 seq 817, from 1.500.001 to 1.765.118, length: 265.118"				
ORIGIN	Query Match 59.6%; Score 280.8; DB 6; Length 265118; Best Local Similarity 75.0%; Pred. No. 1.4e-70; Matches 351; Conservative 0; Mismatches 117; Indels 0; Gaps 0;				
Qy	1	ATGCTACTTCCAGACTGGAAATCAGAAAAGAAATACCTTATAGAGCCATTTTCTGAAGAA	60		
Db	262045	ATGCTCTTCCAGACTGGAAATCAGAAAAGAGATTTTAAATAGCCATTTCTCAGAGGAA	262104		
Qy	61	TGCTCCAAACCCAGCAGGTATACCTCAGAGTGGGCAGAGAGCGTTTTCGTTAAGGGGAAA	120		
Db	262105	TGCTTTCACACGCTGGCTACGACTTAAGGTAGGCAAGGAGCTTACATTCAGGGARAG	262164		
Qy	121	TAAATCGAGTGGAAAGGAAAGTGGTTATTCCTCCAAAGGAATACGCTTAAATC	180		
Db	262165	TTCATAGATGTTGAAAAGAGGGGCAAAAGTCATAATACCTCCAAAAGAAATATGCCCTAATA	262224		
Qy	181	CTNACCTCGAGAGGATAAAGTTGCCGACGATGTTATCGGGGATATGAAGATAAGGAGC	240		
Db	262225	CTGACCCCTAGAGAGGATAAAGCTTCCAGATGATATTTAGGAGACATGAAGATAAGGAGC	262284		
Qy	241	AGTTTAGCAAGAGAGGGGTATTGGTTCTTTTTCGTTGGGTTGACCCAGGATGGGATGGA	300		
Db	262285	AGTCTAGCTAGGGAAGGTGTTCTAGGCTCTTTTGCATGGGTAGACCCGGGATGGGACGT	262344		
Qy	301	AACCTTACACTAATGCTCTACAAATGCCTCAATGAACCTGTGCGAATTAAGATATGGAGAG	360		
Db	262345	AATCTAACCTTGATGCTTTTACAAATGCATCGAAAAGGAGGTAAATTTTAAGGTACAAAGAG	262404		
Qy	361	AGATTTGTGCAGATCGCATTTATAAGGCTAGAGGGTCGGCAGAGAACCCCTTACAGAGGA	420		
Db	262405	AGGTTTGTCCAGATAGCCTTCTTAAAGGCTTGAGGCTCCGCCCAAGAAATCCATACAGGGGC	262464		
Qy	421	AACATCAGGGGAGCAGCAGGTTAGCGTTTTCAAAGAGAAAGAACTC	468		
Db	262465	AACTATCAAGGAAGCAGGAGATAGTCTCTCAAAGAGATCTTAACTC	262512		
RESULT 15	BD133227/c				
LOCUS	BD133227	129 bp	DNA	linear	PAT 18-SEP-2002

DEFINITION	Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying them.				
ACCESSION	BD133227				
VERSION	BD133227.1	GI:23228172			
KEYWORDS	JP 2002505572-A/23.				
SOURCE	unidentified				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 129)				
AUTHORS	Hogrefe,H. and Hansen,C.J.				
TITLE	Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying				
JOURNAL	Patent: JP 2002505572-A 23 19-FEB-2002; STRATAGENE				
COMMENT	OS Unidentified PN JP 2002505572-A/23 PD 19-FEB-2002 PF 20-MAR-1998 JP 1998545828 PR 21-MAR-1997 US 08/842774, 24-OCT-1997 US 08/957709 PI PC C12P19/34,C12Q1/68,C12N9/12 HOLLY HOGREFE,CONNIE J HANSEN CC C12P19/34,C12Q1/68,C12N9/12 CC Strandedness: Single; CC Topology: Linear; CC Polymerase enhancing factor (PEF) extracts, PEF protein CC complexes, CC isolated PEF proteins, and methods for purifying and CC identifying them PH Key Location/Qualifiers FT source 1..129 /organism='Unidentified'. FT source Location/Qualifiers 1..129 /organism="unidentified" /mol_type="genomic DNA" /db_xref="taxon:32644"				
ORIGIN	Query Match 20.5%; Score 96.4; DB 6; Length 129; Best Local Similarity 99.0%; Pred. No. 5.6e-17; Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0				
Qy	1	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAAGAA	60		
Db	98	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAAGAA	99		
Qy	61	TCGCTCCAAACCCAGCAGGTTATGACCTCAGAGTGGGCAG	98		
Db	38	TGCGTCCNACCAGCAGGTTATGACCTCAGAGTGGGCAG	1		
Search completed: July 2, 2005, 07:12:21					
Job time : 2578 secs					

DEFINITION	Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying them.									
ACCESSION	BD133227									
VERSION	BD133227.1 GI:23228172									
KEYWORDS	JP 2002505572-A/23.									
SOURCE	unidentified									
ORGANISM	unidentified									
REFERENCE	1 (bases 1 to 129)									
AUTHORS	Hogrefe, H. and Hansen, C.J.									
TITLE	Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying									
JOURNAL	Patent: JP 2002505572-A 23 19-FEB-2002;									
	STRATAGENE									
COMMENT	OS	Unidentified								
	PN	JP 2002505572-A/23								
	PD	19-FEB-2002								
	PF	20-MAR-1998 JP 1998545828								
	PR	21-MAR-1997 US 08/822774, 24-OCT-1997 US 08/957709 P1								
	PC	HOLLY HOGREFE, CONNIE J HANSEN C12P19/34, C12Q1/68, C12N9/12								
	CC	Strandedness: Single;								
	CC	Topology: Linear;								
	CC	Polymerase enhancing factor (PEF) extracts, PEF protein CC complexes,								
	CC	isolated PEF proteins, and methods for purifying and CC identifying them								
FEATURES	FH	Key		Location/Qualifiers						
	FT	source		1..129						
	FT			/organism='Unidentified'.						
				Location/Qualifiers						
				1..129						
				/organism='unidentified'						
				/mol_type='genomic DNA'						
				/db_xref='taxon:32644'						
ORIGIN										
	Query Match	20.5%; Score 96.4; DB 6; Length 129;								
	Best Local Similarity	99.0%; Pred. No. 5.6e-17;								
	Matches 97; Conservative	0; Mismatches 1; Indels 0; Gaps 0;								
Qy	1	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACCTTATAGAGCCATTTCTGAAGAA								60
Db	98	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACCTTATAGAGCCATTTCTGAAGAA								39
Qy	61	TCGCTCCCAACCAAGCAGGTTATGACCTCAGAGTGGGCAG								98
Db	38	TGGCTCCCAACCAAGCAGGTTATGACCTCAGAGTGGGCAG								1

Search completed: July 2, 2005, 07:12:21
Job time : 2578 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:36:20 ; Search time 30.9767 Seconds
(without alignments)
1942.879 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 156
Sequence: 1 MLLPDWKIRKELIEPFSEE.....PYRGNVQSTRLAFSKKKL 156

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1722976 seqs, 385795295 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	5.1	194	15 US-10-425-114-67700	Sequence 67700, A
2	8	5.1	459	17 US-10-732-923-20236	Sequence 20236, A
3	8	5.1	459	17 US-10-732-923-20237	Sequence 20237, A
4	8	5.1	490	15 US-10-369-493-15077	Sequence 15077, A
5	8	5.1	499	15 US-10-369-493-11671	Sequence 11671, A
6	8	5.1	499	15 US-10-369-493-14468	Sequence 14468, A
7	7	4.5	45	10 US-09-764-891-3318	Sequence 3318, Ap
8	7	4.5	66	16 US-10-437-963-187764	Sequence 187764, A
9	7	4.5	71	16 US-10-425-115-360816	Sequence 360816, A
10	7	4.5	174	16 US-10-767-701-56940	Sequence 56940, A
11	7	4.5	198	16 US-10-437-963-184042	Sequence 184042, A

12	7	4.5	215	17 US-10-732-923-21372	Sequence 21372, A
13	7	4.5	220	16 US-10-437-963-147587	Sequence 147587, A
14	7	4.5	237	10 US-09-461-580A-13	Sequence 13, Appl
15	7	4.5	241	15 US-10-425-114-52925	Sequence 52925, A
16	7	4.5	250	11 US-09-809-665A-83	Sequence 83, Appl
17	7	4.5	250	17 US-10-854-299-83	Sequence 83, Appl
18	7	4.5	288	14 US-10-128-714-3054	Sequence 3054, Ap
19	7	4.5	288	14 US-10-128-714-8054	Sequence 8054, Ap
20	7	4.5	299	14 US-10-219-810-36	Sequence 36, Appl
21	7	4.5	330	16 US-10-437-963-173910	Sequence 173910, A
22	7	4.5	345	14 US-10-279-029-95	Sequence 22, Appl
23	7	4.5	345	14 US-10-219-810-22	Sequence 22, Appl
24	7	4.5	399	15 US-10-369-493-2575	Sequence 2575, Ap
25	7	4.5	429	15 US-10-369-493-18715	Sequence 18715, A
26	7	4.5	448	17 US-10-732-923-20265	Sequence 20265, A
27	7	4.5	466	14 US-10-156-761-9303	Sequence 9303, Ap
28	7	4.5	501	16 US-10-767-701-44737	Sequence 44737, A
29	7	4.5	518	15 US-10-282-122A-70654	Sequence 70654, A
30	7	4.5	592	13 US-10-047-593-2	Sequence 2, Appli
31	7	4.5	592	13 US-10-047-593-4	Sequence 4, Appli
32	7	4.5	607	14 US-10-318-780-10	Sequence 10, Appl
33	7	4.5	609	14 US-10-318-780-11	Sequence 11, Appl
34	7	4.5	609	18 US-10-220-665B-18	Sequence 18, Appl
35	7	4.5	661	15 US-10-369-493-905	Sequence 905, App
36	7	4.5	678	14 US-10-156-761-11079	Sequence 11079, A
37	7	4.5	706	16 US-10-437-963-155249	Sequence 155249, A
38	7	4.5	874	17 US-10-732-923-8334	Sequence 8334, Ap
39	7	4.5	1017	16 US-10-408-765A-158	Sequence 158, App
40	7	4.5	1084	17 US-10-472-928-2452	Sequence 2452, Ap
41	7	4.5	1106	15 US-10-282-122A-45050	Sequence 45050, A
42	7	4.5	1121	9 US-09-370-711-28	Sequence 28, Appl
43	7	4.5	1514	16 US-10-408-765A-1309	Sequence 1309, Ap
44	7	4.5	1688	16 US-10-437-963-188596	Sequence 188596, A
45	7	4.5	2187	16 US-10-437-963-146169	Sequence 146169, A

ALIGNMENTS

RESULT 1
US-10-425-114-67700
; Sequence 67700, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 67700
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3592-094-F1_FU1.pep
US-10-425-114-67700

Query Match 5.1%; Score 8; DB 15; Length 194;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 DLRVGRE 35
DB 139 DLRVGRE 146


```
RESULT 2
US-10-732-923-20236
; Sequence 20236, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 20236
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-20236
Query Match 5.1%; Score 8; DB 17; Length 459;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTLERIKL 68
Db 212 LTLERIKL 219

RESULT 3
US-10-732-923-20237
; Sequence 20237, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 20237
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-20237
Query Match 5.1%; Score 8; DB 17; Length 459;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTLERIKL 68
Db 212 LTLERIKL 219

RESULT 4
US-10-369-493-15077
; Sequence 15077, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15077
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-15077
Query Match 5.1%; Score 8; DB 15; Length 499;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AGYDLRVG 32
Db 454 AGYDLRVG 461

RESULT 5
US-10-369-493-11671
; Sequence 11671, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11671
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11671
Query Match 5.1%; Score 8; DB 15; Length 499;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AGYDLRVG 32
Db 454 AGYDLRVG 461

RESULT 6
US-10-369-493-14468
; Sequence 14468, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14468
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14468
```



```
Query Match          5.1%; Score 8; DB 15; Length 499;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      25 AGYDLRVG 32
      |||||
Db      454 AGYDLRVG 461

RESULT 7
US-09-764-891-3318
; Sequence 3318, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3318
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-3318

Query Match          4.5%; Score 7; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SLAREGV 87
      |||||
Db      29 SLAREGV 35

RESULT 8
US-10-437-963-187764
; Sequence 187764, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 187764
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_84435C.1.pep
US-10-437-963-187764

Query Match          4.5%; Score 7; DB 16; Length 66;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SLAREGV 87
      |||||
Db      5 SLAREGV 11

Query Match          5.1%; Score 8; DB 15; Length 499;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      25 AGYDLRVG 32
      |||||
Db      454 AGYDLRVG 461

RESULT 9
US-10-425-115-360816
; Sequence 360816, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 360816
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_92246C.1.pep
US-10-425-115-360816

Query Match          4.5%; Score 7; DB 16; Length 71;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      128 IRLEGPA 134
      |||||
Db      1 IRLEGPA 7

RESULT 10
US-10-767-701-56940
; Sequence 56940, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 56940
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(174)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: 30951232.pep
US-10-767-701-56940

Query Match          4.5%; Score 7; DB 16; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      57 YALILTL 63
      |||||
Db      94 YALILTL 100

RESULT 11
US-10-437-963-184042
; Sequence 184042, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 184042
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(198)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_81073C.1.pap
US-10-437-963-184042

Query Match 4.5%; Score 7; DB 16; Length 198;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RLEGPAR 135
|||||
DB 181 RLEGPAR 187

RESULT 12
US-10-732-923-21372
; Sequence 21372, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 21372
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Sulfolobus solfataricus
US-10-732-923-21372

Query Match 4.5%; Score 7; DB 17; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 RIKLPDD 71
|||||
DB 24 RIKLPDD 30

RESULT 13
US-10-437-963-147587
; Sequence 147587, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147587
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(220)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_48101C.1.pap
US-10-437-963-147587

Query Match 4.5%; Score 7; DB 16; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LRVGREA 35
|||||
DB 103 LRVGREA 109

RESULT 14
US-09-461-580A-13
; Sequence 13, Application US/09461580A
; Publication No. US20030207325A1
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; TITLE OF INVENTION: LIFESPAN
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-461-580A-13

Query Match 4.5%; Score 7; DB 10; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 EGPARNP 137
|||||
DB 43 EGPARNP 49

RESULT 15
US-10-425-114-52925
; Sequence 52925, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 52925
LENGTH: 241
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3060-025-H5_FLI.pep
US-10-425-114-52925

Query Match 4.5%; Score 7; DB 15; Length 241;
Best Local Similarity 100.0%; Fred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LRVGREA 35
|||
Db 187 LRVGREA 193

Search completed: July 2, 2005, 12:40:44
Job time : 30.9767 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:31:19 ; Search time 33.7674 Seconds
(without alignments)
1786.769 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 156
Sequence: 1 MLPLDKIRKEILIEPFSE.....PYRGNVQSTRFLAFSKKKL 156

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	156	100.0	156	2 AAW72847	Aaw72847 Polymerase
2	22	14.1	154	4 AAB96424	Aab96424 Putative
3	14	9.0	15	2 AAW72845	Aaw72845 Polymerase
4	11	7.1	14	2 AAW72849	Aaw72849 Pyrococcus
5	8	5.1	8	2 AAW72846	Aaw72846 Polymerase
6	8	5.1	490	8 ADS26044	ADS26044 Bacterial
7	8	5.1	499	8 ADS22638	ADS22638 Bacterial
8	8	5.1	499	8 ADS25435	ADS25435 Bacterial
9	7	4.5	15	2 AAW72870	Aaw72870 Polymerase
10	7	4.5	15	4 AAUI8829	AAUI8829 Novel pro
11	7	4.5	45	4 AAM94660	Aam94660 Human rep
12	7	4.5	74	5 ABP07719	ABP07719 Human ORF
13	7	4.5	83	4 AAUS8795	AAUS8795 Propionib
14	7	4.5	83	6 ABMS5314	ABMS5314 Propionib
15	7	4.5	102	6 ABP75762	ABP75762 Human sec
16	7	4.5	109	4 AAO06916	AAO06916 Human pol
17	7	4.5	153	3 AAB14140	Aab14140 Bordetell
18	7	4.5	162	2 AAR96207	Aar96207 Attachmen
19	7	4.5	164	3 AAG46155	Aag46155 Arabidops
20	7	4.5	164	4 AAG82692	Aag82692 S. epider
21	7	4.5	167	5 ABP39663	ABP39663 Staphyloc
22	7	4.5	167	8 ADS05751	ADS05751 Staphyloc
23	7	4.5	171	3 AAG46154	Aag46154 Arabidops
24	7	4.5	178	6 ABP98199	ABP98199 Amino aci
25	7	4.5	180	4 ABB58191	Abb58191 Drosophil

26	7	4.5	182	6 ABP98200	Abp98200 Amino aci
27	7	4.5	186	7 ABO68216	AbO68216 Pseudomon
28	7	4.5	198	3 AAG46153	Aag46153 Arabidops
29	7	4.5	240	6 ADA48536	Ada48536 Rice prot
30	7	4.5	250	3 AAB44561	Aab44561 Virulence
31	7	4.5	250	5 ABP54513	ABP54513 Pasteurel
32	7	4.5	288	6 ABJ25996	Abj25996 Aspergill
33	7	4.5	288	6 ABJ25396	Abj25396 Aspergill
34	7	4.5	299	6 ABP96392	ABP96392 Gossypium
35	7	4.5	345	6 ABP96378	ABP96378 G. hirsut
36	7	4.5	345	7 ADD19145	Add19145 Gossypium
37	7	4.5	345	7 ABO64328	ABO64328 Klebsiell
38	7	4.5	362	3 AAY70024	Aay70024 Human Pro
39	7	4.5	367	7 ABO60852	ABO60852 Klebsiell
40	7	4.5	399	8 ADNI19922	Adni19922 Bacterial
41	7	4.5	413	4 AAU29772	Aau29772 Novel hum
42	7	4.5	419	8 ABM80267	ABM80267 Tumour-as
43	7	4.5	429	8 ADS30682	ADS30682 Bacterial
44	7	4.5	443	5 ABP38922	ABP38922 Staphyloc
45	7	4.5	443	8 ADS06651	ADs06651 Staphyloc

ALIGNMENTS

RESULT 1
AAW72847
ID AAW72847 standard; protein; 156 AA.
XX
AC AAW72847;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 (dUTPase) component.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Peptide 2..15
FT /note= "N-terminal peptide used to generate primers"
XX
PN W09842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
XX
PR 21-OCT-1997; 97US-00957709.
XX
PA (STRA-) STRATAGENE.
XX
Hogrefe H, Hansen CJ;
WPI: 1998-542284/46.
N-PSDB; AAV63860.
XX
PT Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX
PS Claim 17; Page 43; 161pp; English.
XX
CC This is the amino acid sequence of the P45 component of the polymerase
CC enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. The sequence is
CC predicted from a DNA sequence (see AAV63860) obtained from genomic DNA by
CC PCR. P45 and P50 (see AAW72844) are the predominant components of PEF,
CC which acts to enhance the activity of P. furiosus DNA polymerase. P45
CC functions as a dUTPase, and can be used to enhance nucleic acid

CC replication, polymerisation or PCR reactions. The invention provides
CC novel extracts, proteins and complexes that improve the polymerisation
CC activity of nucleic acid polymerases, as well as DNA constructs and
CC antibodies. Also included are methods for identifying compositions with
CC PEF activity, for purifying and using these compositions, and specific
CC extracts, proteins and complexes that function to enhance polymerase
CC activity. Nucleic acid polymerase reactions can be enhanced (claimed) by
CC mixing a nucleic acid template, at least 1 polymerase and a composition
CC having polymerase enhancing activity. Kits are provided for replicating
CC nucleic acids, for site-directed mutagenesis, for nucleic acid sequencing
CC or amplification (preferably PCR or RT-PCR). (Updated on 17-OCT-2003 to
CC standardise OS field)
XX
XX
SQ Sequence 156 AA;

Query Match 100.0%; Score 156; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.1e-149;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILIEPFSESLOPAGYDLRVGREAFVKGKLDVKEGKVIPPREYALI 60
DB 1 MLLPDWKIRKEILIEPFSESLOPAGYDLRVGREAFVKGKLDVKEGKVIPPREYALI 60
QY 61 LTLERIKLPDDVGMGMKIRSSLAKEGVIGSFVWDPGWDGNLTMLYNASNEPVELRYGE 120
DB 61 LTLERIKLPDDVGMGMKIRSSLAKEGVIGSFVWDPGWDGNLTMLYNASNEPVELRYGE 120
QY 121 RFVQIAFIRLEGPARNPYRGNGVQSTRLAFSKRKL 156
DB 121 RFVQIAFIRLEGPARNPYRGNGVQSTRLAFSKRKL 156

RESULT 2
AAB96424
ID AAB96424 standard; protein; 154 AA.
XX
AC AAB96424;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi deoxycytidine deaminase.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-00005034.
XX
PR 21-APR-1999; 99FR-00005034.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
XX WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode proteins
PT useful in industry.
PT
XX
PS Claim 7; Page 1118; 1657pp; French.
XX

CC The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAF86431 and AA441223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to

CC 110 degrees centigrade. Note: This patent is in the same patent family as
CC WO200005062, which contains additional sequences as shown in AAB99132-
CC AAB99143, AAH75903-AAH75920 and AAG66436
XX
SQ Sequence 154 AA;

Query Match 14.1%; Score 22; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 9.4e-14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 89 GSFAWVDFGWDGNLTMLYNAS 110
DB 89 GSFAWVDFGWDGNLTMLYNAS 110

RESULT 3
AAW72845
ID AAW72845 standard; peptide; 15 AA.
XX
AC AAW72845;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 component N-terminal peptide.
XX
XX Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /label= Gly, Ala, Ile
XX
XX WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
XX
PA (STRA-) STRATAGENE.
XX
PI Hogrefe H, Hansen CJ;
XX
XX WPI; 1998-542284/46.
XX
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX
PS Claim 17; Page 33; 161pp; English.

CC This peptide comprises an N-terminal peptide of the P45 component of the
CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It has
CC been obtained from isolated P45; a full-length sequence is provided in
CC AAW72847. PEF, the predominant components of which are P45 and P50 (see
CC AAW72844) proteins, enhances the activity of P. furiosus DNA polymerase,
CC thereby providing replication products of greater length and purity. P45
CC is a dUTPase and possesses polymerase enhancing activity. The invention
CC provides novel extracts, proteins and complexes that improve the
CC polymerisation activity of nucleic acid polymerases. These complexes may
CC include proteins including the P45 N-terminal peptide. Also included are
CC methods for identifying compositions with polymerase enhancing activity,
CC for purifying and using these compositions, and specific extracts,
CC proteins and complexes that function to enhance polymerase activity.
CC Nucleic acid polymerase reactions can be enhanced (claimed) by mixing a
CC nucleic acid template, at least 1 polymerase and a composition having
CC polymerase enhancing activity. Kits are provided for replicating nucleic
CC acids. The kits can be used in site-directed mutagenesis, nucleic acid

CC sequencing or amplification (preferably PCR or RT-PCR). (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX
 SQ Sequence 15 AA;

Query Match 9.0%; Score 14; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.4e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPDWKIRKEILIE 15
 DB 2 LLPDWKIRKEILIE 15
 |||||

RESULT 4
 AAW72849
 ID AAW72849 standard; peptide; 14 AA.
 XX
 AC AAW72849;

XX 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX

DE Pyrococcus furiosus P45 dUTPase uridine-binding motif.

XX Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
 KW amplification; sequencing; replication.

XX Pyrococcus furiosus; strain DSM 3638.

XX WO9842860-A1.

XX 01-OCT-1998.

XX 20-MAR-1998; 98WO-US005497.

XX 21-MAR-1997; 97US-00822774.

XX 24-OCT-1997; 97US-00957709.

XX (STRA-) STRATAGENE.

XX Hogrefe H, Hansen CJ;

XX WPI; 1998-542284/46.

XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.

XX Claim 71; Page 47; 161pp; English.

XX This is the uridine-binding motif of the P45 component (see AAW72847) of
 CC the polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638.
 CC P45 has been identified as a dUTPase that has polymerase enhancing
 CC activity. Sequences are provided (see AAW72849-57) of the uridine-binding
 CC motifs of dUTPases and dCTP deaminases of P. furiosus, Methanococcus
 CC jannaschii, Desulfurolobus ambivalens, Escherichia coli, yeast, human and
 CC herpesvirus; a consensus (see AAW72848) is also provided. A claimed
 CC method of enhancing a nucleic acid polymerase reaction comprises
 CC performing the reaction in the presence of one or more of the following:
 CC a PEF, a dUTPase, a protein that turns-over dUTP and a protein having one
 CC or more of the sequences provided in AAW72848-57. A claimed protein
 CC having PEF activity comprises one or more of sequences given in AAW72848-
 CC 57. Kits are provided for replicating nucleic acids, for site-directed
 CC mutagenesis, for nucleic acid sequencing or for amplification (preferably
 CC PCR or RT-PCR). (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 14 AA;

Query Match 7.1%; Score 11; DB 2; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 FAWVDFGWDGN 101
 DB 1 FAWVDFGWDGN 11
 |||||

RESULT 5
 AAW72846
 ID AAW72846 standard; peptide; 8 AA.
 XX

XX AAW72846;

XX 17-OCT-2003 (revised)

DT 01-MAR-1999 (first entry)

XX Polymerase enhancing factor P45 component N-terminal peptide.

XX Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
 KW amplification; sequencing; replication.

XX Pyrococcus furiosus; strain DSM 3638.

XX WO9842860-A1.

XX 01-OCT-1998.

XX 20-MAR-1998; 98WO-US005497.

XX 21-MAR-1997; 97US-00822774.

XX 24-OCT-1997; 97US-00957709.

XX (STRA-) STRATAGENE.

XX Hogrefe H, Hansen CJ;

XX WPI; 1998-542284/46.

XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.

XX Claim 17; Page 33; 161pp; English.

XX This peptide comprises an N-terminal peptide of the P45 component of the
 CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It has
 CC been obtained from isolated P45. PCR primers (see AAV63861-64) based on
 CC the peptide were used to amplify P45 DNA (see AAV63860). A full-length
 CC P45 sequence is provided in AAW72847. PEF, a predominant component of
 CC which is P45 protein, enhances the activity of P. furiosus DNA
 CC polymerase, thereby providing replication products of greater length and
 CC purity. P45 is a dUTPase and possesses polymerase enhancing activity. The
 CC invention provides novel extracts, proteins and complexes that improve
 CC the polymerisation activity of nucleic acid polymerases. These complexes
 CC may include proteins including the P45 N-terminal peptide. Also included
 CC are methods for identifying compositions with polymerase enhancing
 CC activity, for purifying and using these compositions, and specific
 CC extracts, proteins and complexes that function to enhance polymerase
 CC activity. Nucleic acid polymerase reactions can be enhanced (claimed) by
 CC mixing a nucleic acid template, at least 1 polymerase and a composition
 CC having polymerase enhancing activity. Kits are provided for replicating
 CC nucleic acids. The kits can be used in site-directed mutagenesis, nucleic
 CC acid sequencing or amplification (preferably PCR or RT-PCR). (Updated on
 CC 17-OCT-2003 to standardise OS field)

XX Sequence 8 AA;

Query Match 5.1%; Score 8; DB 2; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1.8e-06;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PDWKIRKE 11

DB 1 PDWKIRKE 8

```
RESULT 6
ADS26044
ID ADS26044 standard; protein; 490 AA.
XX AC
XX ADS26044;
XX AC
XX 02-DEC-2004 (first entry)
XX DT
XX DE Bacterial polypeptide #15077.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
XX US2003233675-A1.
XX PN
XX PD 18-DEC-2003.
XX PF
XX PF 20-FEB-2003; 2003US-00369493.
XX PR
XX PR 21-FEB-2002; 2002US-0360039P.
XX
XX (CAOY/) CAO Y.
XX PA (HINK/) HINKLE G J.
XX PA (SLAT/) SLATER S C.
XX PA (CHEN/) CHEN X.
XX PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX PI WPI; 2004-061375/06.
XX DR
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 15077; 122pp; English.
XX PS
XX
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress
XX condition, improved lignin production or improved galactomannan
XX production. This sequence represents a bacterial polypeptide used in the
XX scope of the invention. Note: The sequence data for this patent did not
XX form part of the printed specification but was obtained in electronic
XX format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 490 AA;
XX
Query Match 5.1%; Score 8; DB 8; Length 490;
```

Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AGYDLRVG 32
| | | | | | | |
Db 445 AGYDLRVG 452

RESULT 7
ADS22638
ID ADS22638 standard; protein; 499 AA.

XX AC ADS22638;
XX DT 02-DEC-2004 (first entry)
XX DE Bacterial polypeptide #11671.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.

XX OS Bacteria.
XX
XX US2003233675-A1.

XX PD 18-DEC-2003.
XX PF 20-FEB-2003; 2003US-00369493.
XX PR 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.
XX PA (HINK/) HINKLE G J.
XX PA (SLAT/) SLATER S C.
XX PA (CHEN/) CHEN X.
XX PA (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX PI WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 11671; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress
XX condition, improved lignin production or improved galactomannan
XX production. This sequence represents a bacterial polypeptide used in the

CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 499 AA;

Query Match 5.1%; Score 8; DB 8; Length 499;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AGYDLRVG 32

|||||||

Db 454 AGYDLRVG 461

RESULT 8

ADS25435

ID ADS25435 standard; protein; 499 AA.

XX

AC ADS25435;

XX

DT 02-DEC-2004 (first entry)

XX

DE Bacterial polypeptide #14468.

XX

KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.

XX

OS Bacteria.

XX

PN US2003233675-A1.

XX

PD 18-DEC-2003.

XX

PF 20-FEB-2003; 2003US-00369493.

XX

PR 21-FEB-2002; 2002US-0360039P.

XX

PA (CAOY/) CAO Y.

PA (HINK/) HINKLE G J.

PA (SLAT/) SLATER S C.

PA (CHEN/) CHEN X.

PA (GOLD/) GOLDMAN B S.

XX

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX

DR WPI; 2004-061375/06.

XX

XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.

XX

PS Claim 1; SEQ ID NO 14468; 122pp; English.

XX

CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification

CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 499 AA;

Query Match 5.1%; Score 8; DB 8; Length 499;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AGYDLRVG 32

|||||||

Db 454 AGYDLRVG 461

RESULT 9

AAW72870

ID AAW72870 standard; peptide; 15 AA.

XX

AC AAW72870;

XX

DT 17-OCT-2003 (revised)

DT 01-MAR-1999 (first entry)

XX

DE Polymerase enhancing factor P45 protein N-terminal peptide.

XX

KW Polymerase enhancing factor; PEF; DNA polymerase; dUTPase; PCR;

KW amplification; sequencing; replication.

XX

OS Pyrococcus furiosus; strain DSM 3638.

XX

FH Key Location/Qualifiers

FT Misc-difference 2 /label= Leu, Tyr

FT Misc-difference 3 /label= Leu, Val

FT Misc-difference 4 /label= Arg, Pro

FT

PN WO9842860-A1.

XX

PD 01-OCT-1998.

XX

PF 20-MAR-1998; 98WO-US005497.

XX

PR 21-MAR-1997; 97US-00822774.

XX

PR 24-OCT-1997; 97US-00957709.

XX

PA (STRA-) STRATAGENE.

XX

PI Hogrefe H, Hansen CJ;

XX

DR WPI; 1998-542284/46.

XX

PT Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.

XX

PS Example 5; Page 33; 161pp; English.

XX

CC This is an N-terminal peptide of the P45 component of the polymerase
CC enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. PEF, the
CC predominant components of which are P50 (see AAW72844) and P45 (see
CC AAW72847), enhances the activity of P. furiosus DNA polymerase. P45
CC functions as a dUTPase. The invention provides novel extracts, proteins
CC and complexes that improve the polymerisation activity of nucleic acid

CC polymerases. These can be used to improve nucleic acid replication.
CC polymerisation and amplification (especially in PCR or RT-PCR). (Updated
CC on 17-OCT-2003 to standardise OS field)

SQ Sequence 15 AA;

Query Match 4.5%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 RKEILLIE 15
|||||||
Db 9 RKEILLIE 15

RESULT 10

AAU18829
ID AAU18829 standard; protein; 45 AA.

XX AC AAU18829;

XX DT 21-NOV-2001 (first entry)

XX DE Novel prostate gland antigen, Seq ID No 128.

XX Human; neurotropic; neuroprotective; cytostatic; antiparkinsonian;
KW antianemic; dermatological; immunosuppressive; antiinflammatory;
KW antihemorrhagic; antirheumatic; virucide; hepatotropic; nephrotropic;
KW osteopathic; prostate gland; prostatitis; adenocarcinoma; hair loss;
KW prostatitis; placental; adenocarcinoma; benign prostatic hypertrophy;
KW hyperplasia; carcinoma; prostate neoplastic disorder; skin aging;
KW reproductive system disorder; autoimmune disorder; urinary system;
KW systemic lupus erythematosus; rheumatoid arthritis; cardiovascular;
KW blood-related disorder; hyperproliferative disorder; respiratory;
KW neurological disorder; endocrine disorder; inflammatory disorder;
KW liver disorder; wound healing; food preservative.

XX OS Homo sapiens.

XX PN WO200155447-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001330.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 14-JUL-2000; 2000US-0217496P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234221P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237032P.
PR 02-OCT-2000; 2000US-0237033P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-476223/51.

DR N-PSDB; AAS30270.

XX Novel isolated prostate gland related polypeptide useful for diagnosis
PT and treatment of disorders of prostate such as prostatodystonia,
PT prostatitis, benign prostatic hypertrophy and malacoplakia.

XX Claim 1; SEQ ID NO 128; 512pp; English.

XX The invention relates to novel isolated prostate gland related nucleic
CC acids (I) and polypeptides (II). (I) and (II) are useful for diagnosis,
CC prognosis, prevention, and/or treatment of diseases and/or disorders of
CC the prostate such as acute non-bacterial prostatitis, chronic non-
CC bacterial prostatitis, acute bacterial prostatitis, prostatodystonia,
CC prostatitis, granulomatous prostatitis, malacoplakia, benign prostatic
CC hypertrophy or hyperplasia, and prostate neoplastic disorders, including
CC adenocarcinomas, transitional cell carcinomas, ductal carcinomas, and
CC squamous cell carcinomas. (I), (II) and antibody to (II) are useful for
CC diagnosing and treating reproductive system disorders (Paget's disease),
CC autoimmune disorders (systemic lupus erythematosus, rheumatoid
CC arthritis), blood-related disorders (sickle cell anaemia),
CC hyperproliferative disorders, urinary system disorders
CC (glomerulonephritis), cardiovascular disorders (arrhythmias), respiratory
CC disorders, musculoskeletal system disorders, neural activity and
CC neurological disorders (Alzheimer's disease and Parkinson's disease),
CC endocrine disorders (Addison's disease), gastrointestinal disorders
CC (inflammatory disorders), liver disorders (biliary liver cirrhosis),
CC pancreatic and gall bladder disorders, disorders of the large intestine,
CC developmental and inherited disorders, diseases at the cellular level,
CC and wound healing and epithelial cell proliferation. (I) or (II) is
CC useful to prevent skin aging, for preventing hair loss, to maintain

Query Match 4.5%; Score 7; DB 4; Length 45;

Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SLAREGV 87
|||||
DB 29 SLAREGV 35

RESULT 11

AAM94660

ID AAM94660 standard; protein; 45 AA.

XX AC AAM94660;

XX DT 21-NOV-2001 (first entry)

XX DE Human reproductive system related antigen SEQ ID NO: 3318.

XX KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy.

XX OS Homo sapiens.

XX PN WO200155320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001339.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 07-JUL-2000; 2000US-0216880P.

XX PR 11-JUL-2000; 2000US-0217487P.

XX PR 11-JUL-2000; 2000US-0217496P.

XX PR 14-JUL-2000; 2000US-0218290P.

XX PR 26-JUL-2000; 2000US-0220963P.

XX PR 26-JUL-2000; 2000US-0220964P.

XX PR 14-AUG-2000; 2000US-0224518P.

XX PR 14-AUG-2000; 2000US-0224519P.

XX PR 14-AUG-2000; 2000US-0225213P.

XX PR 14-AUG-2000; 2000US-0225214P.

XX PR 14-AUG-2000; 2000US-0225266P.

XX PR 14-AUG-2000; 2000US-0225267P.

XX PR 14-AUG-2000; 2000US-0225268P.

XX PR 14-AUG-2000; 2000US-0225270P.

XX PR 14-AUG-2000; 2000US-0225447P.

XX PR 14-AUG-2000; 2000US-0225757P.

XX PR 14-AUG-2000; 2000US-0225758P.

XX PR 14-AUG-2000; 2000US-0225759P.

XX PR 18-AUG-2000; 2000US-0226279P.

XX PR 22-AUG-2000; 2000US-0226681P.

XX PR 22-AUG-2000; 2000US-0226868P.

XX PR 23-AUG-2000; 2000US-0227009P.

XX PR 30-AUG-2000; 2000US-0228924P.

XX PR 01-SEP-2000; 2000US-0229287P.

XX PR 01-SEP-2000; 2000US-0229343P.

XX PR 01-SEP-2000; 2000US-0229344P.

XX PR 01-SEP-2000; 2000US-0229345P.

XX PR 05-SEP-2000; 2000US-0229509P.

XX PR 05-SEP-2000; 2000US-0229513P.

XX PR 06-SEP-2000; 2000US-0230437P.

XX PR 06-SEP-2000; 2000US-0230438P.

```
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0232423P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244517P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX N-PSDB; AAL00630.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX Claim 11; SEQ ID NO 3318; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a protein of the invention
XX
XX SQ Sequence 45 AA;
XX
XX Query Match 4.5%; Score 7; DB 4; Length 45;
XX Best Local Similarity 100.0%; Pred. No. 46;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SLAREGV 87
Db 29 SLAREGV 35
|||||

RESULT 12
ABP07719
ID ABP07719 standard; protein; 74 AA.
XX
AC ABP07719;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:15420.
XX
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX immune deficiency; immune disorder; infectious disease;
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX myasthenia gravis.
XX
XX Homo sapiens.
XX
XX WO200192523-A2.
XX
XX 06-DEC-2001.
XX
```

PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
XX WPI; 2002-106308/14.
DR N-PSDB; ABN23471.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
XX Disclosure; SEQ ID NO 15420; 1037pp; English.
XX
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification)). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 74 AA;
Query Match 4.5%; Score 7; DB 5; Length 74;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 GYDLRVG 32
Db 31 GYDLRVG 37
|||||
RESULT 13
AAU58795
ID AAU58795 standard; protein; 83 AA.
XX
AC AAU58795;
XX
XX 13-FEB-2002 (first entry)
DT
DE Propionibacterium acnes immunogenic protein #19691.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
XX WO200181581-A2.
PN
XX

PD 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US012865.
XX
XX 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
DR N-PSDB; AAS59594.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX Example 1; SEQ ID NO 19990; 1069pp; English.
PS
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 83 AA;
Query Match 4.5%; Score 7; DB 4; Length 83;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 EESLQPA 25
Db 53 EESLQPA 59
|||||
RESULT 14
ABM55314
ID ABM55314 standard; protein; 83 AA.
XX
AC ABM55314;
XX
XX 20-OCT-2003 (first entry)
DT
DE Propionibacterium acnes predicted ORF-encoded polypeptide #19990.
XX
XX Acne vulgaris; antisporrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
XX
XX Propionibacterium acnes.
XX
XX WO2003033515-A1.
PN
XX
XX 24-APR-2003.
PD
XX

PF 11-OCT-2002; 2002WO-US032727.
 PR 15-OCT-2001; 2001US-00978825.
 XX (CORI-) CORIXA CORP.
 XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 PI Barth B, Vallieue-Douglas J;
 XX WPI; 2003-381789/36.
 DR N-PSDB; ACF64523.
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 PT or for stimulating an immune response specific for a P. acnes protein.
 XX Example 1; SEQ ID NO 19990; 1481bp; English.
 XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC polynucleotide of the invention; a polypeptide of the invention; a
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridisation. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open
 CC reading frame) contained within the P. acnes polynucleotides of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 83 AA;
 Query Match 4.5%; Score 7; DB 6; Length 83;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 EESLQPA 25
 Db 53 EESLQPA 59
 RESULT 15
 ABP75762
 ID ABP75762 standard; protein; 102 AA.
 XX AC ABP75762;
 XX 10-FEB-2003 (first entry)
 XX Human secretory polypeptide SPTM SEQ ID NO 946.
 XX Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;
 KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;
 KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;
 KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;
 KW anti-inflammatory; immunosuppressive; neuroprotective; neurotropic;

KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;
 KW antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;
 KW secretory polynucleotide; secretory protein.
 XX Homo sapiens.
 XX WO200283876-A2.
 XX 24-OCT-2002.
 XX 27-MAR-2002; 2002WO-US009921.
 XX 29-MAR-2001; 2001US-0280067P.
 PR 29-MAR-2001; 2001US-0280068P.
 PR 16-MAY-2001; 2001US-0291280P.
 PR 17-MAY-2001; 2001US-0291829P.
 PR 17-MAY-2001; 2001US-0291849P.
 PR 19-JUN-2001; 2001US-0299428P.
 PR 20-JUN-2001; 2001US-0299776P.
 PR 20-JUN-2001; 2001US-0300001P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Dafo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
 PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
 XX WPI; 2003-075543/07.
 DR N-PSDB; ABZ36204.
 XX New human secretory proteins and polynucleotides, useful for diagnosing,
 PT treating or preventing autoimmune/inflammatory disorders (e.g. AIDS),
 PT neurological disorders (e.g. Alzheimer's), or cell proliferations or
 PT cancers.
 XX Claim 27; SEQ ID NO 946; 458pp + Sequence Listing; English.
 XX The invention relates to a secretory polynucleotide (designated sptm)
 CC comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a
 CC naturally occurring polynucleotide sequence at least 90 % identical to
 CC the polynucleotide sequence, a polynucleotide complementary to them or an
 CC RNA equivalent of them. The polypeptide or polynucleotide are useful for
 CC treating, preventing or diagnosing a disease or condition associated with
 CC the expression of functional SPTM. These are particularly useful for
 CC diagnosing, treating or preventing autoimmune/inflammatory disorders
 CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's
 CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,
 CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
 CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,
 CC schizophrenia or amnesia), or cell proliferative disorders (e.g.
 CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,
 CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,
 CC breast, cervix or prostate). The present sequence is one of the SPTM
 CC proteins of the invention (ABP75384-ABP75962). Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 102 AA;
 Query Match 4.5%; Score 7; DB 6; Length 102;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 29 LRVGREA 35
 Db 61 LRVGREA 67
 Search completed: July 2, 2005, 12:36:15
 Job time : 35.7674 secs

244

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:36:20 ; Search time 80.0233 Seconds
(without alignments)
1942.879 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 403
Sequence: 1 MLHHVLIYATSKRLVGKK.....KMKRELARIWDEIEKXLS 403

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1722976 seqs, 385795295 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pcp.*
18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pcp.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pcp.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pcp.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	9	2.2	249	15 US-10-369-493-19428	Sequence 19428, A
2	9	2.2	1185	16 US-10-437-963-173561	Sequence 173561, A
3	9	2.2	2517	16 US-10-437-963-131531	Sequence 131531, A
4	8	2.0	61	16 US-10-767-701-59532	Sequence 59532, A
5	8	2.0	169	17 US-10-732-923-9748	Sequence 9748, Ap
6	8	2.0	205	16 US-10-437-963-162719	Sequence 162719, A
7	8	2.0	221	16 US-10-425-115-191365	Sequence 191365, A
8	8	2.0	224	16 US-10-437-963-175459	Sequence 175459, A
9	8	2.0	282	16 US-10-739-930-7332	Sequence 7332, Ap
10	8	2.0	379	16 US-10-425-115-208551	Sequence 208551, A
11	8	2.0	387	15 US-10-369-493-20327	Sequence 20327, A

12	8	2.0	391	15 US-10-369-493-1260	Sequence 1260, A
13	8	2.0	398	14 US-10-156-761-12051	Sequence 12051, A
14	8	2.0	463	9 US-09-858-075A-2	Sequence 2, Appl1
15	8	2.0	481	15 US-10-335-977-8870	Sequence 8870, Ap
16	8	2.0	486	15 US-10-335-977-8871	Sequence 8871, Ap
17	8	2.0	504	16 US-10-425-115-308208	Sequence 308208, A
18	8	2.0	529	15 US-10-425-114-55078	Sequence 55078, A
19	8	2.0	1142	16 US-10-437-963-204428	Sequence 204428, A
20	8	2.0	1261	17 US-10-732-923-12863	Sequence 12863, A
21	8	2.0	1353	17 US-10-732-923-12887	Sequence 12887, A
22	7	1.7	34	9 US-09-764-887-161	Sequence 161, App
23	7	1.7	34	14 US-10-073-961-161	Sequence 161, App
24	7	1.7	38	16 US-10-425-115-302549	Sequence 302549, A
25	7	1.7	52	15 US-10-424-593-185982	Sequence 185982, A
26	7	1.7	62	15 US-10-424-593-183144	Sequence 183144, A
27	7	1.7	67	16 US-10-425-115-189080	Sequence 189080, A
28	7	1.7	71	15 US-10-424-593-176160	Sequence 176160, A
29	7	1.7	76	15 US-10-425-114-37949	Sequence 37949, A
30	7	1.7	78	16 US-10-425-115-274292	Sequence 274292, A
31	7	1.7	82	15 US-10-424-593-256623	Sequence 256623, A
32	7	1.7	88	16 US-10-425-115-334632	Sequence 334632, A
33	7	1.7	92	15 US-10-424-593-278881	Sequence 278881, A
34	7	1.7	96	16 US-10-437-963-140652	Sequence 140652, A
35	7	1.7	104	11 US-09-864-408A-8856	Sequence 8856, Ap
36	7	1.7	108	14 US-10-029-386-31294	Sequence 31294, A
37	7	1.7	109	15 US-10-104-047-2710	Sequence 2710, Ap
38	7	1.7	114	15 US-10-424-593-264395	Sequence 264395, A
39	7	1.7	114	15 US-10-425-114-39809	Sequence 39809, A
40	7	1.7	117	15 US-10-425-114-68065	Sequence 68065, A
41	7	1.7	118	16 US-10-425-115-321625	Sequence 321625, A
42	7	1.7	119	16 US-10-425-115-193317	Sequence 193317, A
43	7	1.7	122	16 US-10-425-115-260281	Sequence 260281, A
44	7	1.7	127	15 US-10-074-978A-401	Sequence 401, App
45	7	1.7	140	16 US-10-437-963-203014	Sequence 203014, A

ALIGNMENTS

RESULT 1
US-10-369-493-19428
; Sequence 19428, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19428
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-10-369-493-19428

Query Match 2.2%; Score 9; DB 15; Length 249;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 192 LEGKRVLT 200
Db 1 LEGKRVLT 9
RESULT 2

US-10-437-963-173561
; Sequence 173561, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 173561
; LENGTH: 1185
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_71587C.1.pep
US-10-437-963-173561
Query Match 2.2%; Score 9; DB 16; Length 1185;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 385 MKKRELAER 393
Db 164 MKKRELAER 172
RESULT 3
US-10-437-963-131531
; Sequence 131531, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 131531
; LENGTH: 2517
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2517)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3358C.1.pep
US-10-437-963-131531
Query Match 2.2%; Score 9; DB 16; Length 2517;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 385 MKKRELAER 393
Db 163 MKKRELAER 171

RESULT 4
US-10-767-701-59532
; Sequence 59532, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 59532
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 7219797.pep
US-10-767-701-59532
Query Match 2.0%; Score 8; DB 16; Length 61;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 197 VLVTAGAT 204
Db 12 VLVTAGAT 19
RESULT 5
US-10-732-923-9748
; Sequence 9748, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 9748
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Hamamelis virginiana
US-10-732-923-9748
Query Match 2.0%; Score 8; DB 17; Length 169;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 LPTGNPVI 73
Db 52 LPTGNPVI 59
RESULT 6
US-10-437-963-162719
; Sequence 162719, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad

```
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 162719
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_61784C.1.pep
US-10-437-963-162719

Query Match      2.0%; Score 8; DB 16; Length 205;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 GKEVLVTA 201
DB 173 GKEVLVTA 180

RESULT 7
US-10-425-115-191365
; Sequence 191365, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 191365
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_106104C.1.pep
US-10-425-115-191365

Query Match      2.0%; Score 8; DB 16; Length 221;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ERLKLG 159
DB 171 ERLKLG 178

RESULT 8
US-10-437-963-175459
; Sequence 175459, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
```

```
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 175459
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(224)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_73301C.1.pep
US-10-437-963-175459

Query Match      2.0%; Score 8; DB 16; Length 224;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 KKLGVVERFI 162
DB 211 KKLGVVERFI 218

RESULT 9
US-10-739-930-7332
; Sequence 7332, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 7332
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C12642_3.p
US-10-739-930-7332

Query Match      2.0%; Score 8; DB 16; Length 282;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVVTAGAT 204
DB 231 VLVVTAGAT 238

RESULT 10
US-10-425-115-208551
; Sequence 208551, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 208551
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
```

```
; LOCATION: (1)...(379)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MPT4577_121792C.1.pcp
US-10-425-115-208551

Query Match          2.0%; Score 8; DB 16; Length 379;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 LEGKRLV 199
DB 103 LEGKRLV 110
|||||

RESULT 11
US-10-369-493-20327
; Sequence 20327, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20327
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-20327

Query Match          2.0%; Score 8; DB 15; Length 387;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TSKEKLE 342
DB 371 TSKEKLE 378
|||||

RESULT 12
US-10-369-493-1260
; Sequence 1260, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1260
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1260

Query Match          2.0%; Score 8; DB 15; Length 391;
```

```
Best Local Similarity 100.0%; Pred. No. 74;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TSKEKLE 342
DB 371 TSKEKLE 378
|||||

RESULT 13
US-10-156-761-12051
; Sequence 12051, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-282
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12051
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12051

Query Match          2.0%; Score 8; DB 14; Length 398;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVTAGAT 204
DB 99 VLVTAGAT 106
|||||

RESULT 14
US-09-858-075A-2
; Sequence 2, Application US/09858075A
; Patent No. US200202524A1
; GENERAL INFORMATION:
; APPLICANT: Cook, James W.
; APPLICANT: Bulawa, Christine
; TITLE OF INVENTION: ASPERGILLUS N-MYRISTOYL TRANSFERASE GENES AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 07334/089001
; CURRENT APPLICATION NUMBER: US/09/858,075A
; CURRENT FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 09/163,444
; PRIOR FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus N-myrystoyl
US-09-858-075A-2

Query Match          2.0%; Score 8; DB 9; Length 463;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATKSRKLV 17
DB 154 ATKSRKLV 161
|||||
```

RESULT 15
US-10-335-977-8870
; Sequence 8870, Application US/10335977
; Publication No. US2004005279A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8870:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 481 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...481
; SEQUENCE DESCRIPTION: SEQ ID NO: 8870:
US-10-335-977-8870

Query Match 2.0%; Score 8; DB 15; Length 481;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 336 SKEKLIIEE 343
|||||||
Db 281 SKEKLIIEE 288

Search completed: July 2, 2005, 12:40:44
Job time : 80.0233 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model
Run on: July 2, 2005, 12:31:19 ; Search time 87.2326 Seconds
(without alignments)
1786.769 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 403
Sequence: 1 MLHHVLIYATKSRKLVGKK.....MKKRLAARIWDEIEKKLS 403

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_16Dec04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	297	73.7	403	AAW72844	Aaw72844 Polymeras
2	60	14.9	401	AAB96653	Aab96653 Putative
3	30	7.4	406	ADN46559	Adn46559 Thermococ
4	28	6.9	31	AAW72841	Aaw72841 Polymeras
5	24	6.0	24	AAW72863	Aaw72863 Polymeras
6	24	6.0	24	AAW72843	Aaw72843 Polymeras
7	17	4.2	17	AAW72842	Aaw72842 Polymeras
8	13	3.2	35	AAW72860	Aaw72860 Polymeras
9	12	3.0	13	AAW72866	Aaw72866 Polymeras
10	11	2.7	15	AAW72865	Aaw72865 Polymeras
11	11	2.7	15	AAW72864	Aaw72864 Polymeras
12	10	2.5	15	AAW72858	Aaw72858 Polymeras
13	10	2.5	394	ABP80942	Abp80942 N. gonorr
14	10	2.5	399	ABBA47692	Abba47692 Listeria
15	9	2.2	249	ADS30395	Ads30395 Bacterial
16	8	2.0	8	AAW72871	Aaw72871 Polymeras
17	8	2.0	154	ADN46167	Adn46167 Thermococ
18	8	2.0	189	ADA36408	Ada36408 Acinetoba
19	8	2.0	387	ADN41897	Adn41897 Bacterial
20	8	2.0	391	ADN18607	Adn18607 Bacterial
21	8	2.0	434	ABBA48095	Abba48095 Listeria
22	8	2.0	463	AAE04541	Aae04541 Aspergill
23	8	2.0	463	AAE20169	Aae20169 Aspergill
24	8	2.0	481	AAW55465	Aaw55465 H. pylori
25	8	2.0	486	AAW55242	Aaw55242 H. pylori

26	8	2.0	490	3	AAG04881	Aag04881 Arabidops
27	8	2.0	491	3	AAG53291	Aag53291 Arabidops
28	8	2.0	492	3	ADR85982	Adr85982 Aspergill
29	8	2.0	505	3	AAG04880	Aag04880 Arabidops
30	8	2.0	505	8	ADR51287	Adr51287 Anti-biof
31	8	2.0	506	3	AAG53290	Aag53290 Arabidops
32	8	2.0	511	3	AAG04879	Aag04879 Arabidops
33	8	2.0	511	3	AAG53289	Aag53289 Arabidops
34	7	1.7	7	7	AAW72873	Aaw72873 Polymeras
35	7	1.7	9	7	ADD22376	Add22376 HLA-B46 c
36	7	1.7	15	6	ABR39272	AbR39272 N-termina
37	7	1.7	17	2	AAW72862	Aaw72862 Polymeras
38	7	1.7	18	4	AAB92176	Aab92176 Signal tr
39	7	1.7	34	4	AAU91528	Aau91528 Human dig
40	7	1.7	34	4	AAU19974	Aau19974 Human liv
41	7	1.7	34	5	ABP40835	Abp40835 Human liv
42	7	1.7	34	7	ADJ14953	Adj14953 Human liv
43	7	1.7	47	5	AAW47154	Aaw47154 Modular e
44	7	1.7	60	5	ADE31821	Ade31821 Surfactin
45	7	1.7	76	5	AAO21912	Aao21912 BASB204 f

ALIGNMENTS

RESULT 1
AAW72844
ID AAW72844 standard; protein; 403 AA.
XX
AC AAW72844;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P50 component.
XX
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
KW sequencing; replication.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Peptide 1..31
FT Misc-difference 23 /note= "N-terminal peptide used to generate primers"
FT Misc-difference 24 /note= "encoded by NNN"
FT Misc-difference 24 /note= "encoded by NNN"
FT Misc-difference 25 /note= "encoded by NNN"
FT Peptide 275..291
FT Misc-difference 310 /note= "internal peptide used to generate primers"
FT Misc-difference 311 /note= "encoded by CCN"
FT Misc-difference 311 /note= "encoded by NNN"
FT Peptide 353..376
FT /note= "internal peptide used to generate primers"
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1999; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
XX
PA (STRA-) STRATAGENE.
XX
PI Hogrefe H, Hansen CJ;
XX
DR WPI; 1998-542284/46.

DR N-PSDB; AAV63859.
XX
PT Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX
PS Claim 46; Page 36; 161pp; English.
XX
CC This is the amino acid sequence of the P50 component of the polymerase
CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. The sequence is
CC predicted from a DNA sequence (see AAV63859) obtained from genomic DNA by
CC PCR. P50 and P45 (see AAV72847) are the predominant components of PEF,
CC which acts to enhance the activity of P. furiosus DNA polymerase, thereby
CC providing replication products of greater length and purity. P50 is
CC similar in structure to a bacterial flavoprotein. The invention provides
CC novel extracts, proteins and complexes that improve the polymerisation
CC activity of nucleic acid polymerases, as well as DNA constructs and
CC antibodies raised against P45 or P50. Also included are methods for
CC identifying compositions with polymerase enhancing activity, for
CC purifying and using these compositions, and specific extracts, proteins
CC and complexes that function to enhance polymerase activity. Nucleic acid
CC polymerase reactions can be enhanced (claimed) by mixing a nucleic acid
CC template, at least 1 polymerase and a composition having polymerase
CC enhancing activity. Kits are provided for replicating nucleic acids. The
CC kits can be used in site-directed mutagenesis, nucleic acid sequencing or
CC amplification (preferably PCR or RT-PCR). (Updated on 17-OCT-2003 to
CC standardise OS field)
XX
XX Sequence 403 AA;
Query Match 73.7%; Score 297; DB 2; Length 403;
Best Local Similarity 99.8%; Pred. No. 1.7e-294;
Matches 402; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLHHVKLIIVATSKRLVGGKIVXXPGSIAALDVKACGLIRHGAHVHVMSEAATKIIH 60
DB 1 MLHHVKLIIVATSKRLVGGKIVXXPGSIAALDVKACGLIRHGAHVHVMSEAATKIIH 60
QY 61 PYAWNLPNGNVTITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVTT 120
DB 61 PYAWNLPNGNVTITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVTT 120
QY 121 VTTTAPPHIPIMTAPAMHETMYRHPVIVRENIERLKKLGVEFGPRIEEGRKAVSIDELV 180
DB 121 VTTTAPPHIPIMTAPAMHETMYRHPVIVRENIERLKKLGVEFGPRIEEGRKAVSIDELV 180
QY 181 YRVIKKLHKKTLGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGAVTLIR 240
DB 181 YRVIKKLHKKTLGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGAVTLIR 240
QY 241 TGSKVKAFRIRIKLVKVTVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEKGIKS 300
DB 241 TGSKVKAFRIRIKLVKVTVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEKGIKS 300
QY 301 GRSITIELVXPXPKIIDRIKEIQPNVFLVGFKAETSKEKLIIEGKQIERAKADLVVGN 360
DB 301 GRSITIELVXPXPKIIDRIKEIQPNVFLVGFKAETSKEKLIIEGKQIERAKADLVVGN 360
QY 361 LEAFSGSEENQVVLIGRDFTKELPKMKKRELAIRWDEIEKXLS 403
DB 361 LEAFSGSEENQVVLIGRDFTKELPKMKKRELAIRWDEIEKXLS 403
RESULT 2
AAB96653
ID AAB96653 standard; protein; 401 AA.
XX
XX AAB96653;
XX
DT 29-OCT-2001 (first entry)
DE Putative P. abyssi flavoprotein involved in panthothenate metabolism.
XX

KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS *Pyrococcus abyssi*.
XX FR2792651-Al.
XX
XX 27-OCT-2000.
XX
XX 21-APR-1999; 99FR-00005034.
XX
XX 21-APR-1999; 99FR-00005034.
XX
XX (CNRS) CNRS CENT NAT RECH SCI.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins
XX useful in industry.
XX
XX Claim 7; Page 1403-1404; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of *Pyrococcus*
XX *abyssi* (see AAF8431 and AAH41223-7) and *P. abyssi* proteins. *P. abyssi* is
XX a hyperthermophilic archaeon, which is isolated from deep-sea
XX hydrothermal vents. The present sequence is one such *P. abyssi* protein.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade. Note: This patent is in the same patent family as
XX WO200005082, which contains additional sequences as shown in AAB99132-
XX AAB99143, AAH75903-AAH75920 and AAG66436
XX
XX Sequence 401 AA;
Query Match 14.9%; Score 60; DB 4; Length 401;
Best Local Similarity 100.0%; Pred. No. 3.8e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 80 IEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVTTVTTAFPHIPIMIAPAMHE 139
DB 81 IEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVTTVTTAFPHIPIMIAPAMHE 140
RESULT 3
ADN46559
ID ADN46559 standard; protein; 406 AA.
XX
XX ADN46559;
XX
XX 01-JUL-2004 (first entry)
XX
XX *Thermococcus kodakaraensis* KOD1 protein sequence SeqID437.
XX
XX gene disruption; gene targeting; marker gene; transformation;
XX homologous recombination; hyperthermostable archaeobacterium; KOD1;
XX gene structure; gene function; enzyme activity; medicine;
XX forensic science; food; drug inspection; molecular biology; immunology.
XX
XX *Thermococcus kodakaraensis*.
XX
XX WO2004022736-Al.
XX
XX 18-MAR-2004.
XX
XX 29-AUG-2003; 2003WO-IB003597.
XX
XX 30-AUG-2002; 2002JP-00319011.
XX
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

PI Imanaka T, Atomi H;
 DR WPI; 2004-257583/24.
 XX
 XX Method for disrupting targeted gene in genome of organism particularly
 PT thermostable bacterium and with genome chips for analysis, applicable in
 PT studying gene structure and functions.
 XX
 XX
 XX
 XX Claim 9; SEQ ID NO 437; 598pp; Japanese.
 XX
 CC This invention relates to a novel method for targeting disruption of an
 CC arbitrary gene in a genome of an organism which comprises providing the
 CC whole sequential data of the genome of such organism, selecting at least
 CC 1 arbitrary region in the sequence, providing a vector that contains a
 CC sequence homologous with the selected region and a marker gene,
 CC transformation, and homologous recombination. The genome is preferably
 CC the genome of a hyperthermostable archaeobacterium, particularly
 CC Thermococcus kodakaraensis KOD1. The method is for targeting the
 CC disruption of a gene in the genome of an organism, which is applicable in
 CC studying gene structure and functions as well as enzyme activities of
 CC encoded proteins and useful in medicine, forensic science, food or drug
 CC inspection, molecular biology and immunology. With this method, the
 CC disruption of a gene at an arbitrary position in a genome can be achieved
 CC efficiently and reliably. The present sequence is that of a protein
 CC encoded by the genome of Thermococcus kodakaraensis which was derived
 CC using the method of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 406 AA;
 SQ
 Query Match 7.4%; Score 30; DB 8; Length 406;
 Best Local Similarity 100.0%; Pred. No. 1.8e-21;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 97 LVCPTANTISKIACGIDDPVTVTVTAF 126
 DB 101 LVCPTANTISKIACGIDDPVTVTVTAF 130
 |||||
 RESULT 4
 AAW72841
 ID AAW72841 standard; peptide; 31 AA.
 AC
 AC AAW72841;
 XX
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 XX Polymerase enhancing factor P50 component N-terminal peptide.
 XX
 XX Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 KW
 XX Pyrococcus furiosus; strain DSM 3638.
 OS
 XX Key Location/Qualifiers
 FH Misc-difference 23 /note= "unidentified amino acid"
 FT Misc-difference 24 /note= "unidentified amino acid"
 FT Misc-difference 25 /note= "unidentified amino acid"
 FT
 FT
 XX WO9842860-A1.
 PN
 XX
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 XX
 XX 21-MAR-1997; 97US-00822774.
 PR 24-OCT-1997; 97US-00957709.
 PR

(STRA-) STRATAGENE.
 Hogrefe H, Hansen CJ;
 WPI; 1998-542284/46.
 XX
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 XX Claim 10; Page 91; 161pp; English.
 CC This peptide comprises the N-terminal portion of the P50 component of the
 CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It is
 CC obtained from the translated sequence (see AAW72844) of isolated P50 DNA
 CC (see AAV63859) and corresponds to a chemically determined N-terminal
 CC peptide (see AAW72860). PEF, the predominant components of which are P50
 CC and P45 (see AAW72847) proteins, enhances the activity of P. furiosus DNA
 CC polymerase, thereby providing replication products of greater length and
 CC purity. The invention provides novel extracts, proteins and complexes
 CC that improve the polymerisation activity of nucleic acid polymerases.
 CC These complexes may include proteins including the P50 N-terminal
 CC peptide. Also included are methods for identifying compositions with
 CC polymerase enhancing activity, for purifying and using these
 CC compositions, and specific extracts, proteins and complexes that function
 CC to enhance polymerase activity. Nucleic acid polymerase reactions can be
 CC enhanced (claimed) by mixing a nucleic acid template, at least 1
 CC polymerase and a composition having polymerase enhancing activity. Kits
 CC are provided for replicating nucleic acids. The kits can be used in site-
 CC directed mutagenesis, nucleic acid sequencing or amplification
 CC (preferably PCR or RT-PCR). (Updated on 17-OCT-2003 to standardise OS
 CC field)
 XX
 XX Sequence 31 AA;
 SQ
 Query Match 6.9%; Score 28; DB 2; Length 31;
 Best Local Similarity 100.0%; Pred. No. 1.8e-20;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLHHVKLIYATSKRLVGKKIVXXPGSIAA 31
 DB 1 MLHHVKLIYATSKRLVGKKIVXXPGSIAA 31
 |||||
 RESULT 5
 AAW72863
 ID AAW72863 standard; peptide; 24 AA.
 AC
 AC AAW72863;
 XX
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 XX Polymerase enhancing factor P50 component tryptic peptide.
 XX
 XX Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 KW
 XX Pyrococcus furiosus; strain DSM 3638.
 OS
 XX Key Location/Qualifiers
 FH Misc-difference 1 /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT
 FT
 XX WO9842860-A1.
 PN
 XX
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 XX
 XX 21-MAR-1997; 97US-00822774.
 PR 21-MAR-1997; 97US-00822774.

```

PR 24-OCT-1997; 97US-00957709.
XX (STRA-) STRATAGENE.
XX Hogrefe H, Hansen CJ;
PI WPI; 1998-542284/46.
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX Example 5; Page 32; 161pp; English.
XX This is a tryptic peptide of the P50 component of the polymerase
CC enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It was obtained
CC by tryptic digestion of P50 protein isolated from a PEF complex. PEF, the
CC predominant components of which are P50 (see AAW72844) and P45 (see
CC AAW72847), enhances the activity of P. furiosus DNA polymerase. The
CC invention provides novel extracts, proteins and complexes that improve
CC the polymerisation activity of nucleic acid polymerases. These can be
CC used to improve nucleic acid replication, polymerisation and
CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
CC standardise OS field)
XX Sequence 24 AA;
SQ
Query Match 6.0%; Score 24; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.8e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 353 ADLVVGNLTLEAFGSEENQVVLIGR 376
DB 1 ADLVVGNLTLEAFGSEENQVVLIGR 24
RESULT 6
AAW72843
ID AAW72843 standard; peptide; 24 AA.
XX AAW72843;
XX 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX Polymerase enhancing factor P50 component internal peptide.
XX Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
XX sequencing; replication.
XX Pyrococcus furiosus; strain DSM 3638.
XX WO9842860-A1.
XX 01-OCT-1998.
XX 20-MAR-1998; 98WO-US005497.
XX 21-MAR-1997; 97US-00822774.
XX 24-OCT-1997; 97US-00957709.
XX (STRA-) STRATAGENE.
XX Hogrefe H, Hansen CJ;
XX WPI; 1998-542284/46.
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX Claim 11; Page 91; 161pp; English.
XX
PR 24-OCT-1997; 97US-00957709.
XX (STRA-) STRATAGENE.
XX Hogrefe H, Hansen CJ;
PI WPI; 1998-542284/46.
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX Example 5; Page 32; 161pp; English.
XX This is a tryptic peptide of the P50 component of the polymerase
CC enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It was obtained
CC by tryptic digestion of P50 protein isolated from a PEF complex. PEF, the
CC predominant components of which are P50 (see AAW72844) and P45 (see
CC AAW72847), enhances the activity of P. furiosus DNA polymerase. The
CC invention provides novel extracts, proteins and complexes that improve
CC the polymerisation activity of nucleic acid polymerases. These can be
CC used to improve nucleic acid replication, polymerisation and
CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
CC standardise OS field)
XX Sequence 24 AA;
SQ
Query Match 6.0%; Score 24; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.8e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 353 ADLVVGNLTLEAFGSEENQVVLIGR 376
DB 1 ADLVVGNLTLEAFGSEENQVVLIGR 24
RESULT 7
AAW72842
ID AAW72842 standard; peptide; 17 AA.
XX AAW72842;
XX 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX Polymerase enhancing factor P50 component internal peptide.
XX Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
XX sequencing; replication.
XX Pyrococcus furiosus; strain DSM 3638.
XX WO9842860-A1.
XX 01-OCT-1998.
XX 20-MAR-1998; 98WO-US005497.
XX 21-MAR-1997; 97US-00822774.
XX 24-OCT-1997; 97US-00957709.
XX (STRA-) STRATAGENE.
XX Hogrefe H, Hansen CJ;
XX WPI; 1998-542284/46.
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX Claim 11; Page 91; 161pp; English.
XX This peptide comprises an internal peptide of the P50 component of the
CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It is
CC obtained from the translated sequence (see AAW72844) of isolated P50 DNA
CC peptide (see AAW72862). PEF, the predominant components of which are P50
CC and P45 (see AAW72847) proteins, enhances the activity of P. furiosus DNA
CC polymerase, thereby providing replication products of greater length and
CC purity. The invention provides novel extracts, proteins and complexes
CC that improve the polymerisation activity of nucleic acid polymerases.
CC These complexes may include proteins including the P50 internal peptide.
CC Also included are methods for identifying compositions with polymerase
CC enhancing activity, for purifying and using these compositions, and
CC specific extracts, proteins and complexes that function to enhance
CC polymerase activity. Nucleic acid polymerase reactions can be enhanced
CC (claimed) by mixing a nucleic acid template, at least 1 polymerase and a
CC composition having polymerase enhancing activity. Kits are provided for
CC replicating nucleic acids. The kits can be used in site-directed
CC mutagenesis, nucleic acid sequencing or amplification (preferably PCR or
CC RT-PCR). (Updated on 17-OCT-2003 to standardise OS field)
XX Sequence 24 AA;
SQ
Query Match 6.0%; Score 24; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.8e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 353 ADLVVGNLTLEAFGSEENQVVLIGR 376
DB 1 ADLVVGNLTLEAFGSEENQVVLIGR 24

```

CC and P45 (see AAW72847) proteins, enhances the activity of *P. furiosus* DNA
 CC polymerase, thereby providing replication products of greater length and
 CC purity. The invention provides novel extracts, proteins and complexes
 CC that improve the polymerisation activity of nucleic acid polymerases.
 CC These complexes may include proteins including the P50 internal peptide.
 CC Also included are methods for identifying compositions with polymerase
 CC enhancing activity, for purifying and using these compositions, and
 CC specific extracts, proteins and complexes that function to enhance
 CC polymerase activity. Nucleic acid polymerase reactions can be enhanced
 CC (claimed) by mixing a nucleic acid template, at least 1 polymerase and a
 CC composition having polymerase enhancing activity. Kits are provided for
 CC replicating nucleic acids. The kits can be used in site-directed
 CC mutagenesis, nucleic acid sequencing or amplification (preferably PCR or
 CC RT-PCR). (Updated on 17-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 17 AA;

Query Match 4.2%; Score 17; DB 2; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09; Mismatches 0; Indels 0; Gaps 0;
 Matches 17; Conservative 0;
 QY 275 KYDVVIMAAVSDFRPK 291
 |||||
 Db 1 KYDVVIMAAVSDFRPK 17

RESULT 8
 AAW72860
 ID AAW72860 standard; peptide; 35 AA.
 XX
 AC AAW72860;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Polymerase enhancing factor P50 component N-terminal peptide.
 XX
 KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 OS *Pyrococcus furiosus*; strain DSM 3638.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 14
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 16
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 21
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 29
 FT /note= "any amino acid"
 FT Misc-difference 30
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 32..35
 FT /note= "these residues may each be deleted or substituted
 FT by any amino acid but are tentatively assigned as
 FT indicated"
 XX
 PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 XX
 XX 21-MAR-1997; 97US-00822774.
 XX 24-OCT-1997; 97US-00957709.
 XX
 XX This is an N-terminal peptide of the P50 component of the polymerase

XX (STRA-) STRATAGENE.
 PA Hogrefe H, Hansen CJ;
 PI WPI; 1998-542284/46.
 DR
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 PS Example 5; Page 32; 161pp; English.
 XX
 CC This is an N-terminal peptide of the P50 component of the polymerase
 CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. It was obtained
 CC by N-terminal sequencing of P50 isolated from a PEF complex. PEF, the
 CC predominant components of which are P50 (see AAW72844) and P45 (see
 CC AAW72847), enhances the activity of *P. furiosus* DNA polymerase. The
 CC invention provides novel extracts, proteins and complexes that improve
 CC the polymerisation activity of nucleic acid polymerases. These can be
 CC used to improve nucleic acid replication, polymerisation and
 CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 35 AA;
 Query Match 3.2%; Score 13; DB 2; Length 35;
 Best Local Similarity 100.0%; Pred. No. 4.4e-05;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LHHVKLIYATKSR 14
 |||||
 Db 3 LHHVKLIYATKSR 15
 RESULT 9
 AAW72866
 ID AAW72866 standard; peptide; 13 AA.
 XX
 AC AAW72866;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Polymerase enhancing factor P50 component N-terminal peptide.
 XX
 KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 OS *Pyrococcus furiosus*; strain DSM 3638.
 XX
 PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 XX
 XX 21-MAR-1997; 97US-00822774.
 XX 24-OCT-1997; 97US-00957709.
 XX
 XX (STRA-) STRATAGENE.
 PA Hogrefe H, Hansen CJ;
 PI WPI; 1998-542284/46.
 DR
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 PS Example 5; Page 33; 161pp; English.
 XX
 CC This is an N-terminal peptide of the P50 component of the polymerase

CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. It was obtained
 CC by N-terminal sequencing of P50 isolated from a PEF complex. PEF, the
 CC predominant components of which are P50 (see AAW72844) and P45 (see
 CC AAW72847), enhances the activity of *P. furiosus* DNA polymerase. The
 CC invention provides novel extracts, proteins and complexes that improve
 CC the polymerisation activity of nucleic acid polymerases. These can be
 CC used to improve nucleic acid replication, polymerisation and
 CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)

XX Sequence 13 AA;
 SQ Query Match 3.0%; Score 12; DB 2; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.00019;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLHHVKLIYATK 12
 DB 1 MLHHVKLIYATK 12
 |||||

RESULT 10
 AAW72865
 ID AAW72865 standard; peptide; 15 AA.
 XX AC AAW72865;
 XX DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX DE Polymerase enhancing factor 100 kDa protein N-terminal peptide.
 XX KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX OS *Pyrococcus furiosus*; strain DSM 3638.
 XX FH Key Location/Qualifiers
 FT Misc-difference 1 /label= Gly, Ala, Met
 FT Misc-difference 13 /label= Lys, Leu
 FT /notes= "these residues may each be deleted or substituted
 FT by any amino acid but are assigned as indicated"

XX WO9842860-A1.
 XX 01-OCT-1998.
 XX 20-MAR-1998; 98WO-US005497.
 XX 21-MAR-1997; 97US-00822774.
 XX 24-OCT-1997; 97US-00957709.
 XX (STRA-) STRATAGENE.
 XX Hogrefe H, Hansen CJ;
 XX WPI; 1998-542284/46.
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 XX the polymerisation activity of nucleic acid polymerases, for use in
 XX amplification, sequencing and replication.
 XX Example 5; Page 33; 161pp; English.
 XX This is an N-terminal peptide of a 100 kDa component of the polymerase
 XX enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. PEF, the
 XX predominant components of which are P50 (see AAW72844) and P45 (see
 XX AAW72847), enhances the activity of *P. furiosus* DNA polymerase. The
 XX invention provides novel extracts, proteins and complexes that improve
 XX the polymerisation activity of nucleic acid polymerases. These can be
 XX used to improve nucleic acid replication, polymerisation and

CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)

XX Sequence 15 AA;
 SQ Query Match 2.7%; Score 11; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.0022;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYATK 12
 DB 2 LHHVKLIYATK 12
 |||||

RESULT 11
 AAW72864
 ID AAW72864 standard; peptide; 15 AA.
 XX AC AAW72864;
 XX DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX DE Polymerase enhancing factor 150 kDa protein N-terminal peptide.
 XX KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX OS *Pyrococcus furiosus*; strain DSM 3638.
 XX FH Key Location/Qualifiers
 FT Misc-difference 1 /label= Gly, Ala, Met
 FT Misc-difference 11 /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"

XX WO9842860-A1.
 XX 01-OCT-1998.
 XX 20-MAR-1998; 98WO-US005497.
 XX 21-MAR-1997; 97US-00822774.
 XX 24-OCT-1997; 97US-00957709.
 XX (STRA-) STRATAGENE.
 XX Hogrefe H, Hansen CJ;
 XX WPI; 1998-542284/46.
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 XX the polymerisation activity of nucleic acid polymerases, for use in
 XX amplification, sequencing and replication.
 XX Example 5; Page 33; 161pp; English.
 XX This is an N-terminal peptide of a 150 kDa component of the polymerase
 XX enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. PEF, the
 XX predominant components of which are P50 (see AAW72844) and P45 (see
 XX AAW72847), enhances the activity of *P. furiosus* DNA polymerase. The
 XX invention provides novel extracts, proteins and complexes that improve
 XX the polymerisation activity of nucleic acid polymerases. These can be
 XX used to improve nucleic acid replication, polymerisation and
 XX amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
 XX standardise OS field)

XX Sequence 15 AA;
 SQ Query Match 2.7%; Score 11; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.0022;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYATK 12
DB 2 LHHVKLIYATK 12

RESULT 12
AAW72858
ID AAW72858 standard; peptide; 15 AA.
XX
XX AAW72858;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
XX Polymerase enhancing factor P50 component N-terminal peptide.
XX
XX Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
KW sequencing; replication.
XX
XX Pyrococcus furiosus; strain DSM 3638.
OS
XX
XX Key Location/Qualifiers
FT Misc-difference 1 /note= "any amino acid"
FT Misc-difference 2
FT /note= "this residue may be deleted or substituted by any
FT amino acid but is tentatively assigned as indicated"
FT Misc-difference 13 /note= "any amino acid"
FT Misc-difference 14 /note= "any amino acid"
FT Misc-difference 15 /note= "any amino acid"
FT /note= "this residue may be deleted or substituted by any
FT amino acid but is tentatively assigned as indicated"
XX
XX WO9842860-A1.
XX
XX 01-OCT-1998.
XX
XX 20-MAR-1998; 98WO-US005497.
XX
XX 21-MAR-1997; 97US-00822774.
XX 24-OCT-1997; 97US-00957709.
XX (STRA-) STRATAGENE.
XX
XX Hogrefe H, Hansen CJ;
XX
XX WPI; 1998-542284/46.
XX
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX
XX Example 5; Page 32; 161pp; English.
XX
XX This is an N-terminal peptide of the P50 component of the polymerase
XX enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It was obtained
XX by N-terminal sequencing of P50 isolated from a PEF complex. PEF, the
XX predominant components of which are P50 (see AAW72844) and P45 (see
XX AAW72847), enhances the activity of P. furiosus DNA polymerase. The
XX invention provides novel extracts, proteins and complexes that improve
XX the polymerisation activity of nucleic acid polymerases. These can be
XX used to improve nucleic acid replication, polymerisation and
XX amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
XX standardise OS field)
XX
XX Sequence 15 AA;
SQ

Query Match 2.5%; Score 10; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYAT 11
DB 3 LHHVKLIYAT 12

RESULT 13
ABP80942
ID ABP80942 standard; protein; 394 AA.
XX
XX AC ABP80942;
XX
DT 07-MAR-2003 (first entry)
XX
XX N. gonorrhoeae amino acid sequence SEQ ID 8414.
DE
XX Antibacterial; infection; vaccine; gene therapy.
KW
XX Neisseria gonorrhoeae.
OS
XX WO200279243-A2.
XX
XX 10-OCT-2002.
XX
XX 12-FEB-2002; 2002WO-IB002069.
XX
XX 12-FEB-2001; 2001GB-00003424.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Fontana MR, Pizza M, Massignani V, Monaci E;
XX
XX WPI; 2003-058415/05;
DR N-PSDB; ABZ41912.
XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection.
XX
XX Disclosure; Page 801; 815pp; English.
XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention
XX
XX Sequence 394 AA;
SQ

Query Match 2.5%; Score 10; DB 6; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALA 227
DB 208 SSGKMGVALA 217

RESULT 14
ABB47692
ID ABB47692 standard; protein; 399 AA.
XX
XX AC ABB47692;
XX
DT 05-FEB-2002 (first entry)
XX
XX Listeria monocytogenes protein #396.
XX
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX

OS Listeria monocytogenes.
XX WO200177335-A2.
XX 18-OCT-2001.
XX 11-APR-2001; 2001WO-FR001118.
XX 11-APR-2000; 2000PR-00004629.
XX (INSP) INST PASTEUR.
XX Buchrieser C, Frangeul L, Couve E, Rueniok C, Fsihi H, Dehoux P;
PI Dusurget O, Cherouani F, Nedjari H, Glaeser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablo B, Wehlant J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and related
PT polypeptides.
XX Claim 6; SEQ ID NO 397; 192pp; French.
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 399 AA;
Query Match 2.5%; Score 10; DB 5; Length 399;
Best Local Similarity 100.0%; Pred. No. 0.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 197 VLVTAGATRE 206
Db |||||||
190 VLVTAGATRE 199
RESULT 15
ADS30395
ID ADS30395 standard; protein; 249 AA.
XX
AC ADS30395;
XX
DT 02-DEC-2004 (first entry)
XX Bacterial polypeptide #19428.
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX bacterial polypeptide.
XX Bacteria.
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 19428; 122pp; English.
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX Sequence 249 AA;
Query Match 2.2%; Score 9; DB 8; Length 249;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 192 LEGKRVLVLT 200
Db |||||||
1 LEGKRVLVLT 9
Search completed: July 2, 2005, 12:36:13
Job time : 89.2326 secs